С. А. Айвазян

Байесовский подход в эконометрическом анализе

Предлагаемая в данном номере журнала консультация посвящена так называемому байесовскому подходу в эконометрическом анализе, основанному на субъективно-вероятностном способе операционализации принципа максимального использования (наряду с исходными статистическими данными) априорной информации об исследуемом процессе.

Байесовские методы широко распространены в теории и практике эконометрического анализа и являются обязательной составной частью современных учебных программ магистерского уровня по эконометрике в ведущих университетах мира. Особенно заметные преимущества (по сравнению с классическими методами) с точки зрения точности получаемых статистических выводов они имеют в условиях относительно малых выборок, что весьма характерно для эконометрического моделирования.

1. «Философия» и общая логическая схема байесовского подхода

Пусть в описании рассматриваемой эконометрической модели (закона распределения анализируемой случайной величины, функции регрессии, временнóго ряда, системы одновременных уравнений и т. п.) участвует s-мерный параметр $\Theta = (\theta_1, \theta_2, \ldots, \theta_s)^{\mathrm{T}}$ и нашей задачей является построение наилучшей, в определенном смысле, статистической оценки $\hat{\Theta}$ этого параметра по имеющимся k-мерным наблюдениям $\overline{X}_i = (x_i^{(1)}, x_i^{(2)}, \ldots, x_i^{(k)})^{\mathrm{T}}$, $i = 1, 2, \ldots, n$. Верхний индекс T здесь и в дальнейшем означает операцию транспонирования вектора или матрицы, прописными буквами будут обозначаться *векторные* величины (записываемые как векторы-столбцы), а *строчные* буквы будут использоваться для обозначения *одномерных* (возможных или наблюденных) значений анализируемых случайных величин.

Байесовский подход является одним из возможных способов формализации и операционализации тезиса, в справедливости которого нет видимых причин сомневаться: *степень нашей разумной уверенности в некотором утверждении* (*касающемся*, *например*, *неизвестного численного значения интересующего нас параметра*) *возрастает и корректируется по мере пополнения имеющейся у нас информации относительно исследуемого явления*. Могут быть различные формы интерпретации и подтверждения этого тезиса, в том числе не имеющие отношения к байесовскому подходу. Одна из них выражена, например, в свойстве состоятельности оценки $\hat{\Theta}_n$ неизвестного параметра Θ: чем больше объем выборки n, на основании которой мы строим свою оценку $\hat{\Theta}_n$, тем бóльшей информацией об этом параметре мы располагаем и тем ближе (в смысле сходимости $\hat{\Theta}_n$ к Θ по вероятности) к истине наше заключение.

Специфика *именно байесовского* способа операционализации этого тезиса основана *на двух положениях*.

1) Во-первых, «*степень нашей разумной уверенности*» *в справедливости некоторого утверждения численно выражается в виде вероятности*. Это означает, что вероятность в байесовском подходе выходит за рамки ее интерпретации в терминах условий статистического ансамбля (см. п. В.2.1 в [Айвазян, Мхитарян (2001а)]), *но относится к одной из категорий субъективной школы теории вероятностей*.

2) Во-вторых, статистик при принятии решения использует в качестве исходной информации одновременно информацию двух типов: *априорную и содержащуюся в исходных статистических данных* (см. п. В.3.2 в [Айвазян, Мхитарян (2001а)]). При этом априорная информация предоставлена ему в виде некоторого *априорного распределения вероятностей* анализируемого неизвестного параметра, которое описывает степень его уверенности в том, что этот параметр примет то или иное значение, *еще до начала сбора исходных статистических данных*. По мере же поступления исходных статистических данных статистик уточняет (пересчитывает) это распределение, *переходя от априорного распределения к апостериорному*, используя для этого известную формулу Байеса:

$$P\{A_i|B\} = \frac{P\{A_i\} \cdot P\{B|A_i\}}{\sum_{j=1}^{N} P\{B|A_j\} \cdot P\{A_j\}}, \tag{1}$$

которая определяет правило вычисления условной вероятности события A_i (при условии, что событие B уже имело место) по безусловной вероятности события A_i и условным вероятностям $P\{B|A_j\}$, $j = 1, 2, \ldots, N$. При этом предполагается, что $A_1, A_2, \ldots, A_N$ образуют полную систему событий, а событие B имеет ненулевую вероятность (т. е. $P\{B\} > 0$).

Общая логическая схема байесовского метода оценивания значений параметров представлена на рис. 1.

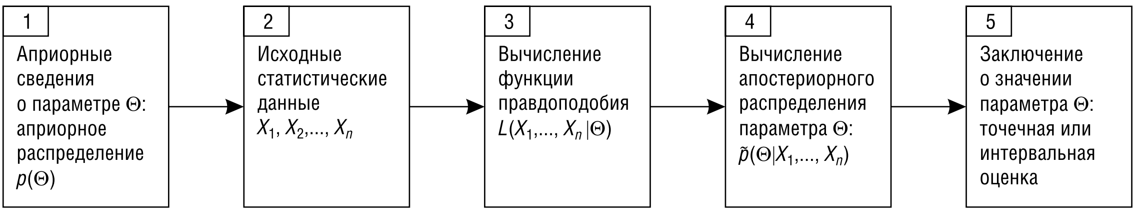

Рис. 1. Общая логическая схема байесовского подхода в статистическом оценивании

Рассмотрим реализацию схемы байесовского оценивания неизвестного параметра.

Априорные сведения о параметре Θ основаны на предыстории функционирования анализируемого процесса (если таковая имеется) и на профессиональных теоретических соображениях о его сущности, специфике, особенностях и т. п. В конечном итоге эти априорные сведения должны быть представлены в виде функции $p(\Theta)$, задающей *априорное распределение параметра* и интерпретируемой как вероятность того, что параметр примет значение, равное Θ, если параметр дискретен, или как функция плотности распределения в точке Θ, если параметр непрерывен по своей природе.

Исходные статистические данные $X_1, X_2, \ldots, X_n$ порождаются в соответствии с законом распределения вероятностей $f(X|\Theta)$, где под $f(X|\Theta)$ понимается значение функции плотности наблюдаемой случайной величины $\xi = (\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(k)})^{\mathrm{T}}$ в точке X, если ξ — непрерывна, или вероятность $P\{\xi = X | \Theta\}$, если ξ дискретна (при условии, что значение неизвестного параметра равно Θ). По умолчанию предполагается, что наблюдения

$$X_1, X_2, \ldots, X_n \tag{2}$$

при фиксированном Θ являются статистически взаимонезависимыми, т. е. образуют *случайную выборку* из анализируемой генеральной совокупности. Так что, получая исходные статистические данные (2), мы к имеющейся априорной информации о параметре (в виде функции $p(\Theta)$) присоединяем соответствующую выборочную (*эмпирическую*) информацию.

Соответственно, **функция правдоподобия** $L(X_1, \ldots, X_n | \Theta)$ (условная, при данном Θ) имеющихся наблюдений (2) определится (с учетом их условной взаимонезависимости) соотношением

$$L(X_1, X_2, \ldots, X_n | \Theta) = f(X_1 | \Theta) \cdot f(X_2 | \Theta) \cdot \ldots \cdot f(X_n | \Theta). \tag{3}$$

Вычисление апостериорного распределения $\tilde{p}(\Theta | X_1, \ldots, X_n)$ осуществляется с помощью формулы Байеса (1) (или ее непрерывного аналога), в которой роль события A_i играет событие, заключающееся в том, что значение оцениваемого параметра равно Θ, а роль условия B — событие, заключающееся в том, что значения n наблюдений, произведенных в анализируемой генеральной совокупности, зафиксированы на уровнях $X_1, X_2, \ldots, X_n$. Соответственно, имеем:

$$\tilde{p}(\Theta | X_1, \ldots, X_n) = \frac{p(\Theta) L(X_1, \ldots, X_n | \Theta)}{\int L(X_1, \ldots, X_n | \Theta) \cdot p(\Theta) d\Theta}. \tag{4}$$

Построение байесовских точечных и интервальных оценок основано на использовании знания апостериорного распределения $\tilde{p}(\Theta | X_1, \ldots, X_n)$, задаваемого соотношением (4). В частности, в качестве байесовских точечных оценок $\hat{\Theta}^{(Б)}$ используют среднее или модальное значение этого распределения, т. е.:

$$\hat{\Theta}^{(Б)}_{(ср)} = \mathbf{E}(\Theta | X_1, \ldots, X_n) = \int \Theta \tilde{p}(\Theta | X_1, \ldots, X_n) d\Theta, \tag{5}$$

$$\hat{\Theta}^{(Б)}_{мод} = \arg\max_{\Theta} \tilde{p}(\Theta | X_1, \ldots, X_n).$$

Отметим, что для определения общего вида апостериорной плотности $\tilde{p}(\Theta | X_1, \ldots, X_n)$ нам достаточно знать только числитель правой части (4), так как знаменатель этого выражения играет роль нормирующего множителя и от Θ не зависит (это существенно упрощает процесс практического построения оценок $\hat{\Theta}^{(Б)}_{ср}$ и $\hat{\Theta}^{(Б)}_{мод}$).

Отметим также одно важное оптимальное свойство оценки $\hat{\Theta}^{(Б)}_{ср}$. Пусть $\hat{\Theta}(X_1, \ldots, X_n)$ — любая оценка параметра Θ. Оказывается, если качество любой оценки $\hat{\Theta}(X_1, \ldots, X_n)$ измерять так называемым *апостериорным байесовским риском*

$$R^{(Б)}(X_1, \ldots, X_n) = \mathbf{E}\{(\hat{\Theta}(X_1, \ldots, X_n) - \Theta)^2 | X_1, \ldots, X_n\} = \int (\hat{\Theta}(X_1, \ldots, X_n) - \Theta)^2 \tilde{p}(\Theta | X_1, \ldots, X_n) d\Theta$$

или его средним (усреднение — по всем возможным выборкам (2)) *значением* $R^{(Б)}_{ср}$, то байесовская оценка (5) является наилучшей и в том и в другом смысле.

Для построения *байесовского доверительного интервала* для параметра Θ необходимо вычислить по формуле (4) функцию $\tilde{p}(\Theta|X_1,\dots X_n)$ апостериорного закона распределения параметра Θ, а затем по заданной доверительной вероятности P_0 определить $100\frac{1+P_0}{2}$ и $100\frac{1-P_0}{2}$%-ные точки этого закона, которые и дают соответственно левый и правый концы искомой интервальной оценки.

Заметим, что байесовский способ оценивания может давать весьма ощутимый выигрыш в точности при *ограниченных* объемах выборок по сравнению с традиционным «частотным» подходом. В процессе же неограниченного роста объема выборки n оба подхода будут давать, в силу их состоятельности, все более похожие результаты.

«Узкие места» или три главных вопроса, возникающие при практической реализации байесовского подхода:

i) как выбрать общий вид (т. е. параметрическое семейство $p(\Theta;D)$) априорного распределения оцениваемого параметра?

ii) как подобрать численные значения D_0 параметров D, определяющие **конкретный** вид априорного распределения при уже сделанном выборе общего вида $p(\Theta;D)$?

iii) как преодолеваются трудности реализации формулы (4) при вычислении апостериорного распределения $\tilde{p}(\Theta|X_1,\dots,X_n)$?

2. Априорные распределения, сопряженные с наблюдаемой генеральной совокупностью (определение и условие существования)

В решении сформулированных выше трех главных вопросов практической реализации байесовского подхода существенную роль играют *распределения, сопряженные с наблюдаемой генеральной совокупностью* (или, что то же, — распределения, сопряженные с функцией правдоподобия $L(X_1,\dots X_n|\Theta)$).

Определение 1. Семейство априорных распределений $G=\{p(\Theta;D)\}$ называется *сопряженным* по отношению к наблюдаемой генеральной совокупности $f(X|\Theta)$ (или по отношению к функции правдоподобия $L(X_1,\dots,X_n|\Theta)$, если и апостериорное распределение $\tilde{p}(\Theta|X_1,\dots,X_n)$, вычисленное по формуле (4), снова принадлежит этому же семейству G.

Другими словами, семейство распределений G сопряжено с $L(X_1,\dots,X_n|\Theta)$, *если оно замкнуто* относительно операции (4) пересчета априорного распределения в апостериорное.

Таким образом, использование в качестве априорных законов распределения вероятностей (з.р.в.) сопряженных по отношению к L плотностей «расшивает» узкое место (iii): поскольку общий вид апостериорного з.р.в. в этом случае известен, остается лишь уметь пересчитывать значения его параметров D при переходе от априорного распределения к апостериорному.

Как мы увидим позже (см. ниже п. 3), использование сопряженных з.р.в. в качестве априорных оказывается в широком классе случаев вполне естественным и оправданным, что позволяет получить ответ и на вопрос (i).

Но всегда ли существует сопряженное по отношению к заданной функции $L(X_1,X_2,\dots X_n|\Theta)$ *распределение, и если оно существует, то как его найти?*

Условие существования сопряженного семейства априорных распределений: если функция правдоподобия $L(X_1, \ldots, X_n | \Theta)$ представима в форме

$$L(X_1,\ldots X_n|\Theta) = v(T_1(X_1,\ldots X_n), \ldots, T_m(X_1,\ldots X_n); \Theta) \cdot \psi(X_1,\ldots X_n), \tag{6}$$

где $T_j(X_1, \ldots X_n)\,(j = 1, 2, \ldots, m)$ и $\psi(X_1, \ldots X_n)$ — некоторые функции от наблюдений $X_1, \ldots, X_n$, не зависящие от параметров Θ, то существует семейство $G = p(\Theta; D)$ априорных распределений, сопряженное с $L(X_1, \ldots, X_n | \Theta)$[1].

Проверка условия существования сопряженного априорного распределения на ряде примеров.

Пример 1.

Анализируемая (наблюдаемая) генеральная совокупность нормальна с неизвестным значением среднего $\mathbf{E}\xi = \theta$ и *известной* дисперсией $\mathbf{D}\xi = \sigma_0^2$ (будем обозначать в дальнейшем подобный факт в форме $\xi \in N_q(\theta; \sigma_0^2)$, где ξ — наблюдаемая случайная величина, а нижний индекс q определяет ее размерность; так что, если $\xi = (\xi^{(1)}, \ldots, \xi^{(k)})^{\mathrm{T}}$ — вектор, то $\xi \in N_k(\Theta; \Sigma_\xi)$ означает, что многомерная случайная величина размерности k распределена нормально с вектором средних значений $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)^{\mathrm{T}}$ и ковариационной матрицей Σ_ξ). В данном примере

$$L(x_1,\ldots,x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\theta)^2} = e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2} \cdot \left(\frac{1}{\sqrt{2\pi}\sigma_0}\right)^n e^{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\bar{x})^2}. \tag{7}$$

Мы видим, что роль функции $v(T(x_1 \ldots x_n); \theta)$ из правой части (6) играет первый сомножитель в правой части (7), причем $m = 1$, $T(x_1, \ldots, x_n) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (достаточная статистика), а следующие за $v(\bar{x}; \theta)$ сомножители правой части (7) от θ не зависят. Следовательно, семейство априорных, *сопряженных с L* распределений существует.

Пример 2.

$\xi \in N_1\left(\theta_1; \frac{1}{\theta_2}\right)$, где и среднее значение $\theta_1 = \mathbf{E}\xi$, и θ_2 — параметр точности $\left(\theta_2 = \frac{1}{\mathbf{D}\xi}\right)$ неизвестны (т. е. $\Theta = (\theta_1, \theta_2)$). Воспользовавшись тем же представлением (7) для функции правдоподобия L, убеждаемся, что $T_1(x_1, \ldots, x_n) = \bar{x}$, $T_2(x_1, \ldots, x_n) = s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ (достаточные статистики), так что в данном случае $m = 2$ и семейство априорных, *сопряженных* (*по отношению к L*) распределений существует.

[1] Функции $T_j(X_1, \ldots, X_n)$, участвующие в представлении (6) (если таковое существует), называются *достаточными статистиками* в задаче статистического оценивания параметров $\Theta = (\theta_1, \theta_2, \ldots, \theta_s)^{\mathrm{T}}$. Размерность m векторной достаточной статистики $T(X_1, \ldots, X_n) = (T_1(X_1, \ldots, X_n), \ldots, T_m(X_1, \ldots, X_n))$ конечна при $n \to \infty$ и зависит от специфики функции L и размерности s оцениваемого параметра Θ. Достаточные статистики играют важную роль в теории и приложениях математической статистики. В частности, они используются в задаче построения наилучших несмещенных оценок в следующей схеме: пусть $\hat{\Theta}$ — некоторая несмещенная оценка параметров Θ и $\mathrm{tr}\,\Sigma < \infty$; тогда $\hat{\Theta}_1 = \mathbf{E}(\hat{\Theta}|T)$ будет снова несмещенной оценкой параметров Θ, причем $\mathrm{tr}\,\Sigma_{\hat{\Theta}_1} \le \mathrm{tr}\,\Sigma_{\hat{\Theta}}$ (под $\Sigma_{\hat{\Theta}}$ понимается ковариационная матрица вектора $\hat{\Theta}$).

Пример 3.

Анализируется *биномиально распределенная* случайная величина $\xi_\theta(M)$ — число «успехов» в серии из M испытаний Бернулли, где θ — неизвестная вероятность «успеха» в одном таком испытании, а M — общее число (известное) испытаний Бернулли в рассматриваемой серии, так что

$$f(x|\theta) = P\{\xi_\theta(M) = x|\theta\} = C_M^x \theta^x (1-\theta)^{M-x}, \;\; x = 0,1,2,\ldots,M.$$

Наблюдаются n таких серий. Тогда

$$L(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^n C_M^{x_i} \theta^{x_i} (1-\theta)^{M-x_i} = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{nM-\sum_{i=1}^n x_i} \prod_{i=1}^n C_M^{x_i},$$

где x_i — число «успехов» в i-й серии.

Поэтому в рамках общего представления (6) в данном случае имеем: $m=1$, $T(x_1,\ldots,x_n) = \sum_{i=1}^n x_i$ — достаточная статистика, что подтверждает существование априорного, сопряженного с L распределения параметра θ.

Пример 4.

Анализируется *отрицательно биномиально распределенная* случайная величина $\xi(\theta;K)$ — число испытаний в схеме Бернулли до K-го появления интересующего нас события, где θ — неизвестная вероятность появления этого события при одном испытании, а K — некоторое *заданное* целое положительное число. Тогда

$$f(x|\theta) = P\{\xi(\theta;K) = x|\theta\} = C_{x-1}^{K-1} \theta^K (1-\theta)^{x-K}, \; x = K, K+1, \ldots,$$

так что

$$L(x_1,\ldots,x_n|\theta) = \prod_{i=1}^n C_{x_i-1}^{K-1} \theta^K (1-\theta)^{x_i-K} = \theta^{Kn} (1-\theta)^{\sum_{i=1}^n x_i - Kn} \cdot \prod_{i=1}^n C_{x_i-1}^{K-1}.$$

Поэтому в рамках общего представления (6) в данном случае имеем: $m=1$, $T(x_1,\ldots,x_n) = \sum_{i=1}^n x_i$ — достаточная статистика, что подтверждает существование априорного, *сопряженного с L* распределения параметра θ.

Пример 5.

В данном примере речь идет об оценивании параметра θ пуассоновского з.р.в., т. е.

$$f(x|\theta) = P\{\xi = x|\theta\} = \frac{\theta^x}{x!} e^{-\theta}, \; x = 0,1,2,\ldots,$$

так что

$$L(x_1, x_2, \ldots, x_n|\theta) = \prod_{i=1}^n \frac{\theta^{x_i}}{x_i!} e^{-\theta} = e^{-n\theta} \cdot \theta^{\sum_{i=1}^n x_i} \cdot \prod_{i=1}^n \left(\frac{1}{x_i!}\right).$$

Сравнивая с общим представлением (6), в данном случае имеем: $m=1$, $T(x_1,\ldots,x_n) = \sum_{i=1}^n x_i$ — достаточная статистика, что подтверждает существование априорного, *сопряженного с L* распределения параметра θ.

Пример 6.

Анализируется *экспоненциально распределенная* (без сдвига) случайная величина с неизвестным значением параметра масштаба θ, т. е.

$$f(x|\theta) = \begin{cases} \theta e^{-\theta x} & \text{при } x \ge 0 \\ 0 & \text{при } x < 0 \end{cases}.$$

Соответственно:

$$L(x_1,\ldots,x_n|\theta) = \begin{cases} \theta^n e^{-\left(\sum\limits_{i=1}^{n} x_i\right)\cdot\theta} & \text{при } x_i \ge 0 \\ 0 & \text{при } x_i < 0 \end{cases}.$$

В рамках общего представления (6) в данном случае имеем: $m = 1$; $T(x_1,\ldots,x_n) = \sum\limits_{i=1}^{n} x_i$ — достаточная статистика, что подтверждает существование априорного, *сопряженного* с L распределения параметра θ.

Пример 7.

Анализируется случайная величина, распределенная *равномерно* на отрезке $[0;\theta]$ при неизвестном значении параметра θ, т. е.

$$f(x|\theta) = \begin{cases} \dfrac{1}{\theta} & \text{при } 0 \le x \le \theta \\ 0 & \text{при } x \notin [0;\theta] \end{cases}.$$

Соответственно:

$$L(x_1,\ldots,x_n|\theta) = \left(\frac{1}{\theta}\right)^n \text{ при } \theta \ge x_{\max}(n) = \max_{1\le i\le n} x_i.$$

Следовательно, в рамках общего представления (6) имеем: $m = 1$; $T(x_1,\ldots,x_n) = x_{\max}(n)$ — достаточная статистика, что подтверждает существование априорного, сопряженного с L распределения параметра θ.

Пример 8.

Анализируется модель распределения Парето с неизвестным значением параметра формы θ, т. е.

$$f(x|\theta) = \begin{cases} \dfrac{\theta x_0^{\theta}}{x^{\theta+1}} & \text{при } x \ge x_0 \\ 0 & \text{при } x < x_0 \end{cases},$$

где пороговое значение x_0 считается заданным.

Соответственно:

$$L(x_1,\ldots,x_n|\theta) = \theta^n x_0^{n\theta} \cdot \left(\prod_{i=1}^{n} x_i\right)^{-(\theta+1)} = \theta^n \left(\frac{g_n}{x_0}\right)^{-n\theta} \cdot g_n^{-n},$$

где $g_n = \left(\prod\limits_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ — среднее геометрическое значение наблюдений $x_1, x_2, \ldots, x_n$ анализируемой случайной величины. Следовательно, обращаясь к (6), имеем $m = 1$, $T(x_1\ldots x_n) = g_n$ —

достаточная статистика, что подтверждает существование априорного, *сопряженного с L* распределения параметра θ.

Пример 9.

Рассмотрим классическую линейную модель множественной регрессии (КЛММР, см., например, [Айвазян (2001), §2.2]) с нормальными, в среднем нулевыми, взаимонезависимыми и гомоскедастичными остатками $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$:

$$Y = \mathbf{X}\Theta + \varepsilon, \tag{8}$$

где $$Y = (y_1, y_2, \ldots, y_n)^{\mathrm{T}} \quad \text{и} \quad \mathbf{X} = \begin{pmatrix} 1 & x_1^{(1)} & \ldots & x_1^{(k)} \\ 1 & x_2^{(1)} & \ldots & x_2^{(k)} \\ \vdots & \ldots & \ldots & \ldots \\ 1 & x_n^{(1)} & \ldots & x_n^{(k)} \end{pmatrix} —$$

наблюденные значения соответственно зависимой (y) и объясняющих ($X = (1, x^{(1)}, \ldots, x^{(k)})^{\mathrm{T}}$) переменных, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\mathrm{T}}$ — случайные регрессионные остатки, а $\Theta = (\theta_0, \theta_1, \ldots, \theta_k)^{\mathrm{T}}$ и $h = (\mathbf{D}\varepsilon_i)^{-1}$ — неизвестные значения параметров модели. Напомним, что значения $\mathbf{X}$, в соответствии с требованиями КЛММР, являются неслучайными и что упомянутые выше свойства регрессионных остатков формулируются в форме условий:

$$\varepsilon \in N_n\left(\mathbf{0}; \frac{1}{h}\mathbf{I}_n\right), \tag{9}$$

где $\mathbf{I}_n$ — единичная матрица размерности n, ковариационная матрица остатков $\Sigma_\varepsilon = \frac{1}{h}\mathbf{I}_n$, а параметр $h = (\mathbf{D}\varepsilon_i)^{-1}$ обычно называют *параметром точности*.

С учетом (8)–(9) функция правдоподобия наблюдений $(\mathbf{X}, Y)$ может быть представлена в форме:

$$L(\mathbf{X}, Y | \Theta; h) = \frac{h^{\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} e^{-\frac{h}{2}(Y-\mathbf{X}\Theta)^{\mathrm{T}}(Y-\mathbf{X}\Theta)}. \tag{10}$$

Но $(Y - \mathbf{X}\Theta)^{\mathrm{T}}(Y - \mathbf{X}\Theta) = (Y - \mathbf{X}\hat{\Theta} + \mathbf{X}\hat{\Theta} - \mathbf{X}\Theta)^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta} + \mathbf{X}\hat{\Theta} - \mathbf{X}\Theta) = [(Y - \mathbf{X}\hat{\Theta}) + \mathbf{X}(\hat{\Theta} - \Theta)]^{\mathrm{T}} \times$ $\times [(Y - \mathbf{X}\hat{\Theta}) + \mathbf{X}(\hat{\Theta} - \Theta)]$, где $\hat{\Theta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y$ — оценка метода наименьших квадратов параметров регрессии Θ.

Поэтому

$$(Y - \mathbf{X}\Theta)^{\mathrm{T}}(Y - \mathbf{X}\Theta) = (Y - \mathbf{X}\hat{\Theta})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta}) + (\hat{\Theta} - \Theta)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\hat{\Theta} - \Theta), \tag{11}$$

так как $(Y - \mathbf{X}\hat{\Theta})^{\mathrm{T}}\mathbf{X}(\hat{\Theta} - \Theta) = [\mathbf{X}(\hat{\Theta} - \Theta)]^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta}) = (Y^{\mathrm{T}}\mathbf{X} - \hat{\Theta}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X})(\hat{\Theta} - \Theta) = [Y^{\mathrm{T}}\mathbf{X} - ((\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y)^{\mathrm{T}} \times$ $\times (\mathbf{X}^{\mathrm{T}}\mathbf{X})](\hat{\Theta} - \Theta) = [Y^{\mathrm{T}}\mathbf{X} - Y^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X})](\hat{\Theta} - \Theta) = 0$.

Возвращаясь к (10) и выражая в (11) сумму квадратов МНК-оцененных остатков $(Y - \mathbf{X}\hat{\Theta})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta})$ через оценку остаточной дисперсии $\hat{\sigma}^2 = \frac{1}{n-k-1}(Y - \mathbf{X}\hat{\Theta})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta})$, имеем:

$$L(\mathbf{X}; Y | \Theta; h) = \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot h^{\frac{n}{2}} e^{-\left(\frac{n-k-1}{2}\hat{\sigma}^2\right)h - \frac{h}{2}(\hat{\Theta}-\Theta)^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{X}(\hat{\Theta}-\Theta)}. \tag{12}$$

Отметим, что $\hat{\sigma}^2$ и $\hat{\Theta}$, в конечном счете, определяются по $Y^{\mathsf{T}}Y$, $\mathbf{X}^{\mathsf{T}}Y$ и $\mathbf{X}^{\mathsf{T}}\mathbf{X}$, так что и в данном случае функция правдоподобия L представима в форме (6), в которой набор достаточных статистик $T(\mathbf{X};Y)$ конечен (при $n \to \infty$) и определяется статистиками $Y^{\mathsf{T}}Y$ и $\mathbf{X}^{\mathsf{T}}Y$. Следовательно, существует априорное распределение параметров Θ и h, сопряженное с L.

3. Генезис априорных сопряженных распределений

Оказывается, для широкого класса наблюдаемых генеральных совокупностей, функции правдоподобия которых допускают представление (6) (т. е. эти генеральные совокупности располагают сопряженным с L априорным распределением своих параметров), справедливо следующее **утверждение о генезисе сопряженных априорных распределений**:

если в байесовском подходе стартовать с априорного распределения, не несущего никакой дополнительной по отношению к имеющимся статистическим данным полезной информации об оцениваемых параметрах, то первый же переход от нее по формуле (4) к апостериорному распределению приведет нас к семейству распределений, сопряженному с наблюдаемой генеральной совокупностью[2].

Именно этот прием поиска априорного распределения, сопряженного с анализируемой функцией правдоподобия, представимой в формуле (6), и предлагается использовать в байесовском подходе.

3.1. Априорные распределения, отражающие «скудость априорных знаний» (САЗ-априорные распределения)

Для математической формализации ситуаций, в которых исследователь не располагает никакой полезной **априорной** информацией о значениях оцениваемого параметра, Джеффрис (см. [Jeffreys (1957)]) предложил следующие два правила выбора соответствующего априорного распределения:

(а) если оцениваемый скалярный параметр θ может (теоретически) принимать значения на конечном интервале $[\theta_{\min}, \theta_{\max}]$ или на бесконечном интервале от $-\infty$ до $+\infty$, то априорную функцию плотности $p(\theta)$ следует считать постоянной на соответствующем интервале;

(б) если же из смысла оцениваемого параметра вытекает, что он может принимать любые положительные значения, то следует считать постоянной на всей числовой прямой $(-\infty; +\infty)$ функцию плотности распределения логарифма от значения параметра, т. е. $p(\ln\theta) = \text{const}$ при $\theta \in (0; +\infty)$.

Будем называть такие априорные распределения «*распределениями, отражающими скудость априорных знаний*» или коротко — «**САЗ-априорными распределениями**». Соответственно, их одномерные функции плотности будем обозначать $p_{\text{САЗ}}(\theta)$, а многомерные — $p_{\text{САЗ}}(\Theta)$.

Тот факт, что для определенных таким образом на бесконечной прямой (полупрямой) априорных распределений *нарушается известное правило нормировки* функции плотности

[2] Строгое доказательство этого утверждения для однопараметрического экспоненциального семейства наблюдаемых генеральных совокупностей см. в [Ghosh et al. (2006), п. 5.1.5]. Однако справедливость этого утверждения подтверждается (непосредственной проверкой) и для весьма широкого класса наблюдаемых генеральных совокупностей, не принадлежащих экспоненциальному семейству.

вероятности (поскольку при этом $\int p_{САЗ}(\theta)\,d\theta \neq 1$, но $\int p_{САЗ}(\theta)\,d\theta = \infty$, где интегрирование проводится по всем возможным значениям θ), не доставляет «технических неудобств»: во-первых, пересчет такой «*несобственной*» априорной функции плотности $p_{САЗ}(\theta)$ в апостериорную по формуле (4) дает уже обычную (*собственную*) функцию плотности $\widetilde{p}_{САЗ}(\theta|X_1,\dots,X_n)$, а во-вторых, при любых сколь угодно больших значениях C плотность

$$p_{САЗ}(\theta) = \begin{cases} \dfrac{1}{2C} & \text{при } \theta \in [-C; +C] \\ 0 & \text{при } \theta \notin [-C; +C] \end{cases}$$

минимизирует энтропийную меру информации $-H = \int_{-C}^{C} p(\theta)\ln p(\theta)\,d\theta$, содержащейся в плотности $p(\theta)$ относительно параметра θ (см., например, [Зельнер (1980), с. 59]). Последнее обстоятельство подтверждает обоснованность использования равномерных распределений $p_{САЗ}(\theta) = \text{const}$ или $p_{САЗ}(\ln\theta) = \text{const}$ в качестве априорных распределений, отражающих скудость априорных знаний (или *САЗ-априорных распределений*).

Замечание 1. Общий вид апостериорного распределения $\widetilde{p}(\theta|X_1,\dots,X_n)$, вычисляемого по формуле (4), определяется, с точностью до нормирующей константы, лишь числителем правой части этой формулы. Поэтому в дальнейшем при анализе равенств, справедливых с точностью до нормирующей константы, мы будем использовать знак ~. Следуя этому правилу, сама формула (4) может быть представлена в виде:

$$\widetilde{p}(\Theta|X_1,\dots,X_n) \sim p(\Theta)\cdot L(X_1,\dots,X_n|\Theta). \tag{4'}$$

Замечание 2. При анализе *многомерных* параметров $\Theta = (\theta_1,\dots,\theta_s)^{\mathrm{T}}$ априорные, в том числе САЗ-априорные, распределения обычно предполагают статистическую независимость компонент $\theta_1,\dots\theta_s$, т. е.

$$p(\Theta) = p(\theta_1)\cdot p(\theta_2)\cdot\ldots\cdot p(\theta_s). \tag{13}$$

И, наконец, в заключение этого пункта определим вид априорной плотности $p(\theta)$ для случая $p(\ln\theta) = \text{const}$, т. е. в ситуации, когда параметр θ может принимать любые, но только положительные значения.

Пусть $F_\theta(y) = P\{\theta < y\}$ — функция распределения параметра θ. Тогда

$$F_\theta(y) = P\{\theta < y\} = P\{\ln\theta < \ln y\} = F_{\ln\theta}(\ln y).$$

Соответственно, функция плотности распределения θ будет:

$$f_\theta(y) = \frac{\partial F_\theta(y)}{\partial y} = \frac{\partial F_{\ln\theta}(\ln y)}{\partial(\ln y)}\cdot\frac{\partial(\ln y)}{\partial y} = f_{\ln\theta}(\ln y)\cdot\frac{1}{y} \sim \frac{1}{y},$$

так как по условию $f_{\ln\theta}(\ln y) = p(\ln\theta) = \text{const}$. Так что *в сокращенной записи* имеем для положительнозначных параметров θ:

$$p_{САЗ}(\theta) \sim \frac{1}{\theta}, \tag{14а}$$

а для параметров θ с возможными значениями, заполняющими всю числовую прямую,

$$p_{САЗ}(\theta) = \text{const}. \tag{14б}$$

3.2. Общий подход к выводу семейства априорных распределений, сопряженных с наблюдаемой генеральной совокупностью

Общий подход к выводу семейства априорных распределений, сопряженных с наблюдаемой генеральной совокупностью, основан на утверждении об их генезисе, сформулированном в начале пункта 3. Из этого утверждения вытекает, в частности, следующая общая схема определения такого семейства.

Шаг 1: проверка условия (6) существования семейства априорных распределений, сопряженных с функцией правдоподобия L для наблюдаемой генеральной совокупности.

Шаг 2: если функция правдоподобия L допускает представление (6) (т. е. если существует семейство сопряженных априорных распределений $p(\Theta;\Lambda)$), то осуществляется *вывод САЗ-апостериорного распределения* $\widetilde{p}_{\text{САЗ}}(\Theta|X_1,\dots,X_n)$ по формуле (4'), т. е.

$$\widetilde{p}_{\text{САЗ}}(\Theta|X_1,\dots,X_n) \sim p_{\text{САЗ}}(\Theta)\cdot L(X_1,\dots,X_n|\Theta). \tag{15}$$

Правая часть соотношения (15) и будет определять общий вид семейства априорных распределений, сопряженных с наблюдаемой генеральной совокупностью, характеризуемой функцией правдоподобия $L(X_1,X_2,\dots,X_n|\Theta)$.

Продемонстрируем реализацию этой общей схемы на рассмотренных выше примерах 1–9. Очевидно, нам остается реализовать лишь шаг 2 из этой схемы, так как шаг 1 уже был реализован выше (см. п. 2).

Пример 1 (*продолжение*).

$\xi\in N_1(\theta;\sigma_0^2)$, где $\theta=\mathbf{E}\xi$ — оцениваемый (неизвестный) параметр, а $\sigma_0^2=\mathbf{D}\xi$ — известное (заданное) значение дисперсии наблюдаемой случайной величины. Ранее было установлено (см. выше, пример 1, формулу (7)), что в этом случае существует семейство сопряженных априорных распределений параметра θ.

Определим $p_{\text{САЗ}}(\theta)=\text{const}$ и с учетом того, что $L(x_1,\dots,x_n|\theta)\sim e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}$ (см. выше, формулу (7)), имеем:

$$\widetilde{p}_{\text{САЗ}}(\theta|x_1,\dots,x_n)\sim p_{\text{САЗ}}(\theta)\cdot L(x_1,\dots,x_n|\theta)\sim e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}.$$

Но правая часть этого соотношения представляет собой (с точностью до нормирующего множителя, не зависящего от θ) плотность нормального распределения со средним значением $\bar{x}$ и дисперсией σ_0^2/n. Следовательно, семейство сопряженных априорных распределений неизвестного среднего значения θ нормально распределенной генеральной совокупности (при известной дисперсии $\sigma_0^2=\mathbf{D}\xi$) *само принадлежит классу нормальных законов распределения*.

Пример 2 (*продолжение*).

$\xi\in N_1\left(\theta;\frac{1}{h}\right)$, где и среднее значение θ, и параметр точности $h=1/\mathbf{D}\xi$ являются неизвестными (т. е. $\Theta=(\theta,h)$). Ранее было установлено (см. выше, пример 2), что в этом случае существует семейство двумерных сопряженных априорных распределений параметра $\Theta=(\theta,h)$.
Определим $p_{\text{САЗ}}(\theta)=\text{const}$ и $p_{\text{САЗ}}(h)\sim\frac{1}{h}$ и с учетом (13) и того, что

$$L(x_1,\ldots,x_n|\theta,h)\sim h^{\frac{n}{2}}e^{-\frac{h}{2}\cdot\sum_{i=1}^{n}(x_i-\theta)^2}=h^{\frac{n}{2}}e^{-\frac{h}{2}\left[n(\theta-\bar{x})^2+\sum_{i=1}^{n}(x_i-\bar{x})^2\right]}\sim(nh)^{\frac{1}{2}}e^{-\frac{nh}{2}(\theta-\bar{x})^2}\cdot h^{\frac{n}{2}-\frac{1}{2}}e^{-\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)h},\qquad(16)$$

имеем:

$$\tilde{p}_{\text{САЗ}}(\theta,h|x_1,\ldots,x_n)\sim p_{\text{САЗ}}(\theta)\cdot p_{\text{САЗ}}(h)\cdot L(x_1,\ldots,x_n|\theta,h)\sim(nh)^{\frac{1}{2}}e^{-\frac{nh}{2}(\theta-\bar{x})^2}\cdot h^{\frac{n-1}{2}-1}e^{-\left(\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)h}.\qquad(17)$$

Но правая часть (17) представляет собой (с точностью до нормирующего множителя, не зависящего от θ и h, см. Приложение 2) *плотность двумерного гамма-нормального распределения*

$$p(\theta,h)\sim(\lambda_0 h)^{\frac{1}{2}}e^{-\frac{\lambda_0 h}{2}(\theta-\theta_0)^2}\cdot h^{\alpha-1}e^{-\beta h}\qquad(18)$$

с параметрами $\lambda_0=n,\theta_0=\bar{x},\alpha=\dfrac{n-1}{2}$ и $\beta=\dfrac{1}{2}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$.

Следовательно, семейство сопряженных априорных распределений двумерного параметра $\Theta=(\theta,h)$, где θ и h соответственно среднее значение и параметр точности наблюдаемой нормальной генеральной совокупности, *принадлежит классу двумерных гамма-нормальных распределений* (18).

Пример 3 (*продолжение*).

Наблюдаемая случайная величина $\xi_\theta(M)$ подчиняется биномиальному з.р.в. с неизвестным значением вероятности «успеха» θ и заданным числом испытаний Бернулли M.

Ранее было установлено (см. выше, пример 3), что существует семейство сопряженных априорных распределений параметра θ.

Определим $p_{\text{САЗ}}(\theta)=1$ для $\theta\in(0;1)$ и с учетом того, что $L(x_1,\ldots,x_n|\theta)\sim\theta^{\sum_{i=1}^{n}x_i}\cdot(1-\theta)^{nM-\sum_{i=1}^{n}x_i}$, имеем:

$$\tilde{p}_{\text{САЗ}}(\theta|x_1,\ldots,x_n)\sim p_{\text{САЗ}}(\theta)\cdot L(x_1,\ldots,x_n|\theta)\sim\theta^{\sum_{i=1}^{n}x_i}\cdot(1-\theta)^{nM-\sum_{i=1}^{n}x_i}.\qquad(19)$$

Но правая часть соотношения (19) представляет собой (с точностью до нормирующего множителя, не зависящего от θ) *плотность бета-распределения*

$$p(\theta)=\frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}\qquad(20)$$

с параметрами $a=\sum\limits_{i=1}^{n}x_i+1$ и $b=nM-\sum\limits_{i=1}^{n}x_i+1$ (участвующая в правой части (20) в выражении нормирующего множителя функция $\Gamma(z)$ — это известная гамма-функция Эйлера, т. е. $\Gamma(z)=\int\limits_0^\infty x^{z-1}e^{-x}\,dx$).

Следовательно, семейство сопряженных априорных распределений параметра θ (вероятности «успеха») наблюдаемой биномиально распределенной генеральной совокупности *принадлежит классу бета-распределений* (20).

Пример 4 (*продолжение*).

Ранее было установлено (см. выше, пример 4), что *отрицательно биномиально распределенная* случайная величина $\xi(\theta;K)$ имеет сопряженное априорное распределение параметра θ — вероятности «успеха» в одном испытании Бернулли. Как и в предыдущем примере, определяем $p_{CA3}(\theta) = 1$(для $\theta \in (0;1)$). Тогда с учетом того, что $L(x_1, \dots, x_n|\theta) \sim \theta^{Kn}(1-\theta)^{\sum_{i=1}^{n} x_i - Kn}$ (см. выше, пример 4), имеем:

$$\widetilde{p}_{CA3}(\theta|x_1,\dots,x_n) \sim p_{CA3}(\theta)\cdot L(x_1,\dots,x_n|\theta) \sim \theta^{Kn}(1-\theta)^{\sum_{i=1}^{n} x_i - Kn}. \qquad (21)$$

Правая часть (21) представляет собой (с точностью до нормирующего множителя, не зависящего от θ) *плотность бета-распределения* (20) с параметрами $a = Kn + 1$ и $b = \sum_{i=1}^{n} x_i - Kn + 1$.

Так что семейство сопряженных априорных распределений параметра θ (вероятности «успеха») наблюдаемой отрицательно биномиально распределенной случайной величины $\xi(\theta; K)$ *принадлежит классу бета-распределений* (20).

Пример 5 (*продолжение*).

Ранее было установлено (см. выше, пример 5), что параметр θ пуассоновского з.р.в. имеет сопряженное априорное распределение. Из смысла параметра θ следует, что он может принимать только положительные значения, поэтому определяем $p_{CA3}(\theta) \sim \frac{1}{\theta}$. Тогда с учетом того, что $L(x_1, \dots, x_n|\theta) \sim \theta^{\sum_{i=1}^{n} x_i}\cdot e^{-n\theta}$ (см. выше, пример 5), имеем:

$$\widetilde{p}_{CA3}(\theta|x_1,\dots,x_n) \sim p_{CA3}(\theta)\cdot L(x_1,\dots,x_n|\theta) \sim \theta^{\sum_{i=1}^{n} x_i - 1}\cdot e^{-n\theta}. \qquad (22)$$

Правая часть (22) представляет собой (с точностью до нормирующего множителя, не зависящего от θ) *плотность гамма-распределения*

$$p(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\cdot\theta^{\alpha-1} e^{-\beta\theta}, \quad \theta > 0, \qquad (23)$$

с параметрами $\alpha = \sum_{i=1}^{n} x_i$ и $\beta = n$.

Следовательно, семейство сопряженных априорных распределений параметра θ наблюдаемой генеральной совокупности *принадлежит классу гамма-распределений* (23).

Пример 6 (*продолжение*).

Ранее было установлено (см. выше, пример 6), что параметр масштаба θ экспоненциального распределения имеет сопряженное априорное распределение. Поскольку $\theta > 0$, определяем $p_{CA3}(\theta) \sim \frac{1}{\theta}$. Тогда с учетом того, что $L(x_1, \dots, x_n|\theta) \sim \theta^{n}\cdot e^{-\left(\sum_{i=1}^{n} x_i\right)\theta}$ (см. выше, пример 6), имеем:

$$\widetilde{p}_{CA3}(\theta|x_1,\dots,x_n) \sim p_{CA3}(\theta)\cdot L(x_1,\dots,x_n|\theta) \sim \theta^{n-1}\cdot e^{-\left(\sum_{i=1}^{n} x_i\right)\theta}. \qquad (24)$$

Правая часть (24) определяет (с точностью до нормирующего множителя, не зависящего от θ) *плотность гамма-распределения (23)* с параметрами $\alpha = n$ и $\beta = \sum_{i=1}^{n} x_i$. Так что семейство сопряженных априорных распределений параметра масштаба θ экспоненциально распределенной генеральной совокупности *принадлежит классу гамма-распределений (23).*

Пример 7 (*продолжение*).

Как мы видели (см. выше, пример 7), и при равномерно распределенной на отрезке $[0;\theta]$ случайной величине неизвестный параметр θ имеет сопряженное априорное распределение. Поскольку параметр θ может принимать любые *положительные* значения, определяем $p_{\text{СА3}}(\theta) \sim \dfrac{1}{\theta}$. Тогда с учетом того, что $L(x_1, \dots, x_n|\theta) = \left(\dfrac{1}{\theta}\right)^n$ (и $x_{\max}(n) = \max\limits_{1 \le i \le n} x_i \le \theta$), имеем:

$$\widetilde{p}_{\text{СА3}}(\theta|x_1,\dots,x_n) \sim \begin{cases} p_{\text{СА3}}(\theta)\cdot L(x_1,\dots,x_n|\theta) \sim \left(\dfrac{1}{\theta}\right)^{n+1} \text{при } \theta \ge x_{\max}(n) \\ 0 \ \text{ при } \theta < x_{\max}(n) \end{cases}. \tag{25}$$

Но правая часть соотношения (25) представляет собой (с точностью до нормирующего множителя, не зависящего от θ) *плотность распределения Парето* вида

$$p(\theta) = \begin{cases} \dfrac{\alpha\theta_{\min}^{\alpha}}{\theta^{\alpha+1}} & \text{при } \theta \ge \theta_{\min} \\ 0 & \text{при } \theta < \theta_{\min} \end{cases} \tag{26}$$

с параметром формы $\alpha = n$ и некоторым параметром сдвига $\theta_{\min} \ge x_{\max}(n)$. Следовательно, семейство сопряженных априорных распределений параметра θ равномерно (на $[0;\theta]$) распределенной случайной величины *принадлежит классу распределений Парето вида (26).*

Пример 8 (*продолжение*).

В данном примере речь идет о *наблюдаемой* генеральной совокупности, подчиняющейся распределению Парето с неизвестным значением параметра формы θ и некоторым заданным значением параметра сдвига x_0 (см. выше, пример 8), так что

$$L(x_1,\dots,x_n|\theta) = \theta^n \left(\frac{g_n}{x_0}\right)^{-n\theta} \cdot g_n^{-n} \sim \theta^n \cdot e^{-\left[n\ln\left(\frac{g_n}{x_0}\right)\right]\cdot\theta},$$

где $g_n = \left(\prod\limits_{i=1}^{n} x_i\right)^{\frac{1}{n}}$. Но тогда САЗ-апостериорная функция плотности распределения параметра θ будет иметь вид $\left(\text{с учетом того, что } p_{\text{СА3}}(\theta) \sim \dfrac{1}{\theta}\right)$:

$$\widetilde{p}_{\text{СА3}}(\theta|x_1,\dots,x_n) \sim p_{\text{СА3}}(\theta)\cdot L(x_1,\dots,x_n|\theta) \sim \theta^{n-1}\cdot e^{-\left[n\ln\left(\frac{g_n}{x_0}\right)\right]\cdot\theta}. \tag{27}$$

Мы видим, что правая часть соотношения (27) определяет (с точностью до нормирующего множителя, не зависящего от параметра θ) плотность гамма-распределения (23) с параметром $\alpha = n$ и параметром $\beta = n\ln\left(\dfrac{g_n}{x_0}\right)$, так что сопряженные априорные распределения параметра формы θ наблюдаемой Парето-распределенной генеральной совокупности *принадлежат семейству гамма-распределений*.

Пример 9 (*продолжение*).

Выше при рассмотрении нормальной классической линейной модели множественной регрессии с неизвестными значениями коэффициентов регрессии $\Theta = (\theta_0, \theta_1, \ldots, \theta_k)^{\mathrm{T}}$ и параметра точности $h = \frac{1}{\sigma^2}$ (где $\sigma^2 = \mathbf{D}\varepsilon_i$) мы убедились в том, что у параметров $\theta_0, \theta_1, \ldots, \theta_k, h$ существуют *сопряженные* априорные распределения. Определим теперь общий вид сопряженного априорного распределения $p(\Theta; h)$ параметров Θ и h. С учетом «Замечания 2» (см. выше) и положительных значений параметра h имеем:

$$p_{\text{СА3}}(\theta_0, \theta_1, \ldots, \theta_k; h) = p_{\text{СА3}}(\theta_0) \cdot p_{\text{СА3}}(\theta_1) \cdot \ldots \cdot p_{\text{СА3}}(\theta_k) \cdot p_{\text{СА3}}(h) \sim \frac{1}{h}.$$

Используя полученное ранее выражение (12) для функции правдоподобия $L(\mathbf{X}; Y | \Theta; h)$, имеем:

$$\tilde{p}_{\text{СА3}}(\Theta; h | \mathbf{X}; Y) \sim p_{\text{СА3}}(\Theta; h) \cdot L(\mathbf{X}; Y | \Theta; h) \sim \frac{1}{h} \cdot h^{\frac{n}{2}} \cdot e^{-\left(\frac{n-k-1}{2}\hat{\sigma}^2\right)h} \cdot e^{-\frac{h}{2}(\hat{\Theta}-\Theta)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\hat{\Theta}-\Theta)} =$$

$$= h^{\frac{n-k-1}{2}-1} \cdot e^{-\left(\frac{n-k-1}{2}\hat{\sigma}^2\right)h} \cdot h^{\frac{k+1}{2}} \cdot e^{-\frac{h}{2}(\hat{\Theta}-\Theta)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\hat{\Theta}-\Theta)}. \quad (28)$$

Но правая часть соотношения (28) определяет (с точностью до нормирующего множителя, не зависящего от Θ и h) так называемое *многомерное гамма-нормальное распределение* с параметром сдвига $\hat{\Theta}$, матрицей точности $(\mathbf{X}^{\mathrm{T}}\mathbf{X})$ и параметрами $\alpha = \frac{n-k-1}{2}$ и $\beta = \frac{n-k-1}{2}\hat{\sigma}^2$ (подробнее о многомерном гамма-нормальном распределении и его свойствах см. в Приложении 2). Напомним, что $\hat{\Theta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y$ — это МНК-оценка параметров регрессии Θ, а $\hat{\sigma}^2 = \frac{1}{n-k-1}(Y - \mathbf{X}\hat{\Theta})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta})$ — оценка остаточной дисперсии σ^2.

Таким образом, *сопряженные* априорные распределения параметров $(\Theta; h)$ нормальной классической линейной множественной регрессии имеют общий вид:

$$p(\Theta; h) \sim h^{\frac{k+1}{2}} |\Lambda_0|^{\frac{1}{2}} e^{-\frac{h}{2}(\Theta-\Theta_0)^{\mathrm{T}}\Lambda_0(\Theta-\Theta_0)} \cdot h^{\alpha-1}e^{-\beta h}, \quad (29)$$

в котором конкретное задание векторного параметра сдвига Θ_0, $(k+1)\times(k+1)$-матрицы точности Λ_0 и скалярных параметров α и β однозначно определяет априорный закон распределения параметров Θ и h (напомним, что $k+1$ — это общее число объясняющих переменных, включая свободный член, в анализируемой модели регрессии).

Очевидно, семейство многомерных гамма-нормальных распределений (29) является многомерным обобщением двумерного гамма-нормального распределения (18).

3.3. Рекомендации по подбору конкретных значений параметров в сопряженных априорных распределениях

Использование в качестве априорных законов распределения вероятностей (з.р.в.), *сопряженных* с наблюдаемой генеральной совокупностью (в ситуациях, когда они существуют), позволяет нам определить их *общий вид*, т. е. задает целое *семейство* априорных распределений $\{p(\Theta; D)\}$. Однако при реализации байесовского подхода мы должны оперировать *конкретным* априорным распределением, что требует знания числовых значений D_0 параметров D, от которых наш априорный з.р.в. зависит. *Как же подбирать эти значения D_0* в каждом конкретном случае? Ниже описывается один из возможных подходов к решению данной задачи.

В широком классе ситуаций можно исходить из того, что нам известны априорные средние значения оцениваемого параметра $\Theta_0 = \mathbf{E}\Theta = (\mathbf{E}\theta_1, \mathbf{E}\theta_2, \ldots, \mathbf{E}\theta_s)^{\mathrm{T}}$ и их среднеквадратические ошибки $\Delta_1 = \sqrt{\mathbf{D}\theta_1}$, $\Delta_2 = \sqrt{\mathbf{D}\theta_2}, \ldots, \Delta_s = \sqrt{\mathbf{D}\theta_s}$. Тогда параметры априорного распределения, как правило, могут быть определены *методом моментов* (в случае многомерного параметра Θ — с учетом «Замечания 2» о статистической независимости компонент вектора Θ в априорном распределении, см. (13)).

Продемонстрируем реализацию этого подхода на рассмотренных выше примерах.

1) Определение параметров априорного гамма-распределения (см. формулу (23) и примеры 5, 6, 8). Как известно (см., например, [Айвазян, Мхитарян (2001)]), среднее значение ($\mathbf{E}\theta$) и дисперсия ($\mathbf{D}\theta$) гамма-распределения выражаются через параметры α и β этого распределения по формулам:

$$\mathbf{E}\theta = \frac{\alpha}{\beta}, \quad \mathbf{D}\theta = \frac{\alpha}{\beta^2}.$$

Подставляя в эти соотношения вместо $\mathbf{E}\theta$ и $\mathbf{D}\theta$ соответственно заданные значения θ_0 и Δ^2, получаем в качестве решений системы из двух уравнений (относительно α и β):

$$\alpha = \frac{\theta_0^2}{\Delta^2}, \quad \beta = \frac{\theta_0}{\Delta^2}. \tag{30}$$

2) Определение параметров априорного бета-распределения (см. формулу (20) и примеры 3 и 4). Используя выражения для среднего и дисперсии бета-рапсределения (см., например, [Айвазян, Мхитарян (2001)]) и решая систему из двух уравнений

$$\begin{cases} \mathbf{E}\theta = \dfrac{a}{a+b} = \theta_0 \\ \mathbf{D}\theta = \dfrac{ab}{(a+b)^2(a+b+1)} = \Delta^2 \end{cases} \tag{31}$$

относительно a и b, получаем:

$$a = \frac{\theta_0^2(1-\theta_0)}{\Delta^2} - \theta_0, \quad b = \left(\frac{\theta_0^2(1-\theta_0)}{\Delta^2} - \theta_0\right)\frac{1-\theta_0}{\theta_0}.$$

3) Определение параметров априорного распределения Парето (см. формулу (26) и пример 7). В данном случае параметр формы α и параметр сдвига $\theta_{\min}$ определяются по заданным значениям $\theta_0 = \mathbf{E}\theta$ и $\Delta^2 = \mathbf{D}\theta$ из системы уравнений

$$\begin{cases} \mathbf{E}\theta = \dfrac{\alpha\theta_{\min}}{\alpha-1} = \theta_0 \\ \mathbf{D}\theta = \dfrac{\alpha\theta_{\min}^2}{(\alpha-1)^2(\alpha-2)} = \Delta^2 \end{cases}. \tag{32}$$

Решение этой системы относительно α и $\theta_{\min}$ дает:

$$\alpha = 1 + \sqrt{1 + \frac{\theta_0^2}{\Delta^2}}, \quad \theta_{\min} = \frac{1}{\alpha}\theta_0 \cdot (\alpha - 1). \tag{32'}$$

4) Определение параметров двумерного гамма-нормального распределения (см. формулу (18) в примере 2). Из свойств двумерного гамма-нормального распределения следует (см. Приложение 2), что частное априорное распределение параметра h есть гамма-распределение с параметрами α и β. Поэтому, воспользовавшись заданными значениями $h_0 = \mathbf{E}h$ и $\Delta_h^2 = \mathbf{D}h$, составляем систему из двух уравнений относительно α и β:

$$\begin{cases} \mathbf{E}h = \dfrac{\alpha}{\beta} = h_0 \\ \mathbf{D}h = \dfrac{\alpha}{\beta^2} = \Delta_h^2 \end{cases}.$$

Получаем решение:

$$\alpha = \frac{h_0^2}{\Delta_h^2} \text{ и } \beta = \frac{h_0}{\Delta_h^2}. \tag{33}$$

Для определения параметра λ_0 и параметра сдвига θ_0 воспользуемся тем, что частное априорное распределение параметра сдвига θ есть обобщенное распределение Стьюдента с 2α степенями свободы, параметром сдвига θ_0 и параметром точности, равным $\lambda_0 \dfrac{\alpha}{\beta}$ (сведения о $t\left(2\alpha \middle| \theta_0; \lambda_0 \dfrac{\alpha}{\beta}\right)$-распределении см. в Приложении 1). Из свойств этого распределения следует, что $\mathbf{E}t\left(2\alpha \middle| \theta_0; \lambda_0 \dfrac{\alpha}{\beta}\right) = \theta_0$ и $\mathbf{D}t\left(2\alpha \middle| \theta_0; \lambda_0 \dfrac{\alpha}{\beta}\right) = \dfrac{\beta}{\lambda_0 \cdot \alpha} \cdot \dfrac{2\alpha}{2\alpha - 2}$ (см. Приложение 1), так что при заданных значениях $\theta_0 = \mathbf{E}\theta$ и $\Delta_\theta^2 = \mathbf{D}\theta$ имеем:

- Значение параметра сдвига в распределении (18) равно θ_0;
- $\Delta_\theta^2 = \dfrac{\beta}{\lambda_0 \cdot \alpha} \cdot \dfrac{\alpha}{\alpha - 1}$, откуда $\lambda_0 = \dfrac{1}{\Delta_\theta^2} \cdot \dfrac{\beta}{\alpha - 1}$ (34)

(напомним, что α и β уже определены соотношениями (33)).

5) Определение параметров многомерного гамма-нормального распределения (см. формулу (29) в примере 9). Воспользуемся свойствами многомерного гамма-нормального распределения (см. Приложение 2). В соответствии с ними:

(i) частное распределение числового сомножителя h является гамма-распределением с параметрами α и β;

(ii) частное распределение параметра Θ есть обобщенное $(k + 1)$-мерное распределение Стьюдента с 2α числом степеней свободы, параметром сдвига Θ_0 и матрицей точности $B = \dfrac{\alpha}{\beta}\Lambda_0$ (мы обозначаем его как $t(2\alpha | \Theta_0; B)$-распределение).

Свойство (i) позволяет (при заданных значениях $h_0 = \mathbf{E}h$ и $\Delta_h^2 = \mathbf{D}h$) определить значения параметров α и β по той же формуле (33).

Свойство (ii), дополненное «Замечанием 2» и правилами вычисления вектора средних значений и ковариационной матрицы $(k + 1)$-мерной случайной величины $t\left(2\alpha \middle| \Theta_0; \dfrac{\alpha}{\beta}\Lambda_0\right)$

(см. Приложение 1), позволяет определить остальные параметры распределения (29) — параметр сдвига Θ_0 и элементы матрицы Λ_0. Действительно:

$$\mathbf{E}\Theta = \mathbf{E}t\left(2\alpha|\Theta_0\,;\frac{\alpha}{\beta}\Lambda_0\right) = \Theta_0 \text{ (задано!)}$$

$$\Sigma_\Theta = \Sigma_{t(2\alpha|\Theta_0;\frac{\alpha}{\beta}\Lambda_0)} = \frac{2\alpha}{2\alpha-2}\left(\frac{\alpha}{\beta}\Lambda_0\right)^{-1} = \begin{pmatrix} \Delta_0^2 & & & 0 \\ & \Delta_1^2 & & \\ & & \ddots & \\ 0 & & & \Delta_k^2 \end{pmatrix}, \tag{35}$$

где Δ_j^2 — заданные значения априорных дисперсий компонент вектора $\Theta = (\theta_0, \theta_1, \ldots, \theta_k)$, $j = 0, 1, \ldots, k$.

Таким образом, векторный параметр сдвига в распределении (29) определяется заданным вектором априорных средних значений Θ_0, а диагональные элементы $\lambda_0^{(j)}$ ($j = 0, 1, \ldots, k$) матрицы Λ_0 определяются из уравнений (35) по формуле:

$$\lambda_0^{(j)} = \frac{1}{\Delta_j^2}\cdot\frac{\beta}{\alpha-1}, \tag{36}$$

где значения α и β определены соотношениями (33).

4. Пересчет значений параметров при переходе от априорного сопряженного распределения к апостериорному

Поскольку, по определению, семейство сопряженных априорных распределений $\{p(\Theta; D)\}$ *замкнуто* относительно операции (4) пересчета априорного распределения в апостериорное, то общий вид апостериорного распределения $\tilde{p}(\Theta|X_1, \ldots, X_n)$ при использовании *сопряженных* априорных распределений нам известен, и нам лишь надо уметь пересчитывать параметры $\tilde{D}(X_1, \ldots, X_n)$ этого апостериорного распределения по заданным параметрам D_0 априорного распределения и имеющимся наблюдениям $X_1, X_2, \ldots, X_n$.

Общая схема такого пересчета следующая. Пусть $\{p(\Theta; D)\}$ — семейство априорных распределений, сопряженных с функцией правдоподобия $L(x_1, \ldots, x_n|\Theta)$ имеющихся у нас наблюдений ($D = (d_1, \ldots, d_q)$ — вектор параметров, от которых зависит сопряженное априорное распределение $p(\Theta; D)$), и пусть D_0 — заданные (известные) значения параметров D в анализируемом случае. Тогда с помощью ряда тождественных преобразований правая часть соотношения

$$\tilde{p}(\Theta|X_1, \ldots, X_n) \sim p(\Theta; D_0)\cdot L(X_1, \ldots, X_n|\Theta) \tag{37}$$

приводится, с точностью до множителей, не зависящих от Θ, к виду $p(\Theta; D(X_1, \ldots, X_n))$, где последняя функция принадлежит семейству $p(\Theta; D)$, а каждая из компонент $d_j(X_1, \ldots, X_n)$ ($j = 1, 2, \ldots, q$) вектора параметров $D(X_1, \ldots, X_n)$ является функцией от D_0 и $\{X_1, X_2, \ldots, X_n\}$.

Продемонстрируем реализацию этой общей схемы на наших примерах (с разной степенью подробности).

Пример 1 (*продолжение*).

В данном примере $L(x_1, x_2, \ldots, x_n|\theta) \sim e^{-\frac{n}{2\sigma_0^2}(\bar{x}-\theta)^2}$ (см. (7)), $p(\theta; D) = \frac{1}{\sqrt{2\pi}\Delta_0} e^{-\frac{(\theta-\theta_0)^2}{2\Delta_0^2}}$ (т. е. $d_1 = \theta_0$, $d_2 = \Delta_0^2$), так что

$$\tilde{p}(\theta|x_1, \ldots, x_n) \sim e^{-\frac{(\theta-d_1)^2}{2d_2}} \cdot e^{-\frac{1}{2\sigma_0^2/n}(\bar{x}-\theta)^2} \sim e^{-\frac{1}{2\tilde{d}_2}(\theta-\tilde{d}_1)^2},$$

где

$$\tilde{d}_1(x_1, \ldots, x_n) = \mathbf{E}(\theta|x_1, \ldots, x_n) = \frac{\frac{1}{\sigma_0^2/n} \cdot \bar{x} + \frac{1}{\Delta_0^2} \cdot \theta_0}{\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2}} \tag{38}$$

и

$$\tilde{d}_2(x_1, \ldots, x_n) = \mathbf{D}(\theta|x_1, \ldots, x_n) = \left(\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2}\right)^{-1}.$$

Необходимые промежуточные выкладки нацелены на выделение полного квадрата разности $(\theta - \tilde{d}_1)^2$ из выражения $\frac{1}{ad_2}(\theta - d_1)^2 + \frac{1}{2\sigma_0^2/n}(\theta - \bar{x})^2$ и не представляют принципиальных трудностей.

Мы видим, что среднее ($\tilde{d}_1$) и дисперсия ($\tilde{d}_2$) апостериорного нормального распределения являются определенным образом средневзвешенными значениями априорных и выборочных соответственно средних и дисперсий.

Пример 2 (*продолжение*).

При реализации общей схемы пересчета априорных параметров в апостериорные в данном случае следует учесть представление функции правдоподобия L в форме (16) (см. выше, пример 2), вид (18) априорной плотности двумерного гамма-нормального распределения (в котором вектор параметров $D_0 = (\lambda_0; \theta_0; \alpha; \beta)$), а также справедливость тождества

$$n(\theta - \bar{x})^2 + \lambda_0(\theta - \theta_0)^2 = (\lambda_0 + n)\left(\theta - \frac{\lambda_0\theta_0 + n\bar{x}}{\lambda_0 + n}\right)^2 + \frac{\lambda_0 n}{\lambda_0 + n}(\theta_0 - \bar{x})^2.$$

Тогда вычисление $\tilde{p}(\theta; h)$ по схеме (37) приводит нас снова к двумерному гамма-нормальному распределению вида (18), но с параметрами

$$\begin{aligned} \tilde{\lambda}_0 &= \lambda_0 + n, \\ \tilde{\theta}_0 &= \frac{n\bar{x} + \lambda_0\theta_0}{n + \lambda_0}, \\ \tilde{\alpha} &= \alpha + \frac{n}{2}, \\ \tilde{\beta} &= \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{(\bar{x} - \theta_0)^2}{2\left(\frac{1}{n} + \frac{1}{\lambda_0}\right)}. \end{aligned} \tag{39}$$

Пример 3 (*продолжение*).

Непосредственная реализация соотношения (37) в данном случае дает:

$$\tilde{p}(\theta|x_1,\ldots,x_n) \sim \theta^{a-1}(1-\theta)^{b-1}\cdot\theta^{\sum_{i=1}^n x_i}(1-\theta)^{nM-\sum_{i=1}^n x_i} = \theta^{a+\sum_{i=1}^n x_i-1}(1-\theta)^{b+nM-\sum_{i=1}^n x_i-1}. \quad (40)$$

Но правая часть (40) определяет (с точностью до нормирующего множителя) снова бета-распределение с параметрами:

$$\tilde{a} = a+\sum_{i=1}^n x_i, \qquad \tilde{b} = b+nM-\sum_{i=1}^n x_i. \quad (41)$$

Пример 4 (*продолжение*).

Подставляя в правую часть соотношения (37)

$$p(\theta;D_0) = p(\theta;a,b) \sim \theta^{a-1}(1-\theta)^{b-1},$$

$$L(x_1,\ldots,x_n|\theta) \sim \theta^{Kn}(1-\theta)^{\sum_{i=1}^n x_i-Kn},$$

имеем:

$$\tilde{p}(\theta|x_1,\ldots,x_n) \sim \theta^{a+Kn-1}(1-\theta)^{b+\sum_{i=1}^n x_i-Kn-1}.$$

Мы видим, что апостериорное распределение параметра (вероятности «успеха») отрицательно-биномиального закона, так же как и априорное, является бета-распределением и что его параметры $\tilde{D} = (\tilde{a}, \tilde{b})$ определяются соотношениями:

$$\tilde{a} = a+Kn, \quad \tilde{b} = b+\sum_{i=1}^n x_i - Kn. \quad (42)$$

Пример 5 (*продолжение*).

Как мы видели ранее, функция правдоподобия наблюдений пуассоновской генеральной совокупности имеет вид:

$$L(x_1,\ldots,x_n|\theta) \sim \theta^{\sum_{i=1}^n x_i} e^{-n\theta}.$$

Так что, используя в качестве априорного распределения $p(\theta;D) = p(\theta;\alpha,\beta)$ параметра θ гамма-распределение (23), имеем:

$$\tilde{p}(\theta|x_1,\ldots,x_n) \sim \theta^{\alpha-1}e^{-\beta\theta}\cdot\theta^{\sum_{i=1}^n x_i}e^{-n\theta} = \theta^{\alpha+\sum_{i=1}^n x_i-1}e^{-(\beta+n)\theta}.$$

Тем самым подтверждается сопряженность априорного гамма-распределения, причем *апостериорное* гамма-распределение определяется параметрами $\tilde{D} = (\tilde{\alpha}, \tilde{\beta})$, где

$$\tilde{\alpha} = \alpha+\sum_{i=1}^n x_i; \quad \tilde{\beta} = \beta+n. \quad (43)$$

Пример 6 (*продолжение*).

Функция правдоподобия экспоненциально распределенных наблюдений (с параметром масштаба θ) имеет вид:

$$L(x_1,\ldots,x_n|\theta) = \theta^n e^{-\left(\sum_{i=1}^n x_i\right)\theta}.$$

Так что при априорном гамма-распределении параметра θ имеем:

$$\widetilde{p}(\theta|x_1,\ldots,x_n)\sim\theta^{\alpha-1}e^{-\beta\theta}\cdot\theta^n e^{-\left(\sum_{i=1}^{n}x_i\right)\theta}=\theta^{\alpha+n-1}e^{-\left(\beta+\sum_{i=1}^{n}x_i\right)\theta}.$$

Мы видим, что апостериорное распределение параметра θ снова подчиняется закону гамма-распределения (23), но с параметрами:

$$\widetilde{\alpha}=\alpha+n,\quad \widetilde{\beta}=\beta+\sum_{i=1}^{n}x_i. \qquad (44)$$

Пример 7 (*продолжение*).

Подставляя в правую часть соотношения (37) функцию правдоподобия *равномерно* распределенных (на отрезке $[0;\theta]$) наблюдений и функцию плотности распределения Парето (26) в качестве априорного распределения $p(\theta;D)=p(\theta;\alpha;\theta_{\min})$, имеем:

$$\widetilde{p}(\theta|x_1,\ldots,x_n)\sim\frac{\alpha\theta_{\min}^{\alpha}}{\theta^{\alpha+1}}\cdot\frac{1}{\theta^n}\quad(\text{при }\theta\ge\max\{\theta_{\min};x_1,x_2,\ldots,x_n\}).$$

Отсюда следует, что апостериорное распределение параметра θ описывается, так же как и априорное, законом Парето (26), но с параметрами:

$$\widetilde{\alpha}=\alpha+n,\quad \widetilde{\theta}_{\min}=\max\{\theta_{\min};x_1,x_2,\ldots,x_n\}. \qquad (45)$$

Пример 8 (*продолжение*).

Как мы видели (см. выше, пример 8), функция правдоподобия Парето-распределенных наблюдений имеет вид:

$$L(x_1,\ldots,x_n|\theta)\sim\theta^n\cdot e^{-\left[n\ln\left(\frac{g_n}{x_0}\right)\right]\cdot\theta}.$$

Подставляя ее в правую часть соотношения (37), а также, в качестве априорного распределения $p(\theta;D_0)$, плотность гамма-распределения (23), имеем:

$$\widetilde{p}(\theta|x_1,\ldots,x_n)\sim\theta^{\alpha-1}e^{-\beta\theta}\cdot\theta^n e^{-\left[n\ln\left(\frac{g_n}{x_0}\right)\right]\theta}=\theta^{\alpha+n-1}e^{-\left(\beta+n\ln\left(\frac{g_n}{x_0}\right)\right)\theta},$$

что определяет гамма-распределение с параметрами:

$$\widetilde{\alpha}=\alpha+n,\quad \widetilde{\beta}=\beta+n\ln\left(\frac{g_n}{x_0}\right), \qquad (46)$$

где $g_n=\left(\prod_{i=1}^{n}x_i\right)^{\frac{1}{n}}$ — среднее геометрическое наблюдений $x_1,\ldots,x_n$, а x_0 — параметр сдвига в анализируемом распределении Парето (его значение считается заданным).

Пример 9 (*продолжение*).

Байесовское оценивание коэффициентов регрессии $\Theta=(\theta_0,\theta_1,\ldots,\theta_k)$ и параметра h в нормальной классической модели множественной регрессии (8)–(9) предполагает использование апостериорного распределения $\widetilde{p}(\Theta;h|\mathbf{X},Y)$ этих параметров, определяемого по схеме (37). Подставляя в правую часть соотношения (37) в качестве априорного многомерное гамма-нормальное распределение (29), а также функцию правдоподобия $L(\mathbf{X},Y|\Theta,h)$ (12), преобразованную к виду

$$L(\mathbf{X},Y|\Theta;h)\sim h^{\frac{n-k-1}{2}}e^{-\left(\frac{n-k-1}{2}\hat{\sigma}^2\right)h}\cdot h^{\frac{k+1}{2}}e^{-\frac{h}{2}(\hat{\Theta}-\Theta)^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{X})(\hat{\Theta}-\Theta)},$$

получаем после ряда тождественных преобразований (см. [Де Грот (1974)]) апостериорную плотность $\widetilde{p}(\Theta;h|\mathbf{X},Y)$ в форме многомерного гамма-нормального распределения (29), параметры которого определяются по параметрам $\Theta_0,\Lambda_0,\alpha$ и β априорного распределения и наблюдениям $(\mathbf{X},Y)$ следующими соотношениями:

$$\begin{cases}\widetilde{\Theta}_0=(\Lambda_0+\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\Lambda_0\Theta_0+\mathbf{X}^{\mathrm{T}}Y) \text{ — параметр сдвига;}\\ \widetilde{\Lambda}_0=\Lambda_0+\mathbf{X}^{\mathrm{T}}\mathbf{X} \text{ — матрица точности;}\\ \left.\begin{array}{l}\widehat{\alpha}=\alpha+\dfrac{n}{2};\\ \widetilde{\beta}=\beta+\dfrac{1}{2}\left[(Y-\mathbf{X}\widetilde{\Theta}_0)^{\mathrm{T}}Y+(\Theta_0-\widetilde{\Theta}_0)^{\mathrm{T}}\Lambda_0\Theta_0\right]\end{array}\right\}\begin{array}{l}\text{параметры частного апостериорного}\\ \text{гамма-распределения параметра}\\ \text{точности } h.\end{array}\end{cases}\qquad(47)$$

5. Примеры задач на точечное и интервальное байесовское оценивание параметров модели

Задача 1. Анализ закона распределения домашних хозяйств определенной социально-экономической страты в заданном регионе страны по величине среднедушевого дохода η.

Мы располагаем следующей информацией об анализируемой генеральной совокупности:

(а) логарифм (натуральный) величины среднедушевого дохода (т. е. $\xi=\ln\eta$) домашних хозяйств рассматриваемой страты данного региона распределен нормально с *неизвестным* средним значением θ и *известной* дисперсией $\sigma_0^2=0{,}28$;

(б) имеются результаты обследования $n=10$ случайно отобранных от анализируемой страты домашних хозяйств по среднедушевому доходу y_i (в нижеследующей таблице даны значения $x_i=\ln y_i$):

i	1	2	3	4	5	6	7	8	9	10
x_i	0,54	1,20	0,36	0,80	0,42	2,10	0,70	0,25	0.90	0,48

(в) из предыстории и опыта обследования домашних хозяйств той же страты в других регионах страны получены априорные значения среднего $\mathbf{E}\theta=\theta_0=0{,}60$ и дисперсии $\mathbf{E}\theta=\Delta_0^2=0{,}03$.

Требуется:

Используя сопряженное априорное распределение параметра θ, получить байесовские точечную и интервальную (с уровнем доверия $P_0=0{,}95$) оценки средней величины логарифма среднедушевого дохода и сравнить их с соответствующими оценками метода максимального правдоподобия.

Решение. Мы уже знаем (см. выше, пример 1), что сопряженное априорное распределение в данном случае существует и является нормальным, причем параметры этого распределения непосредственно заданы ($\mathbf{E}\theta = \theta_0 = 0{,}60$ и $\mathbf{D}\theta = \Delta_0^2 = 0{,}03$). В соответствии с выведенными выше формулами пересчета (см. п. 4, формулы (38)) имеем:

$$\tilde{\theta}_0 = \mathbf{E}(\theta | x_1, ..., x_n) = \frac{\frac{1}{\sigma_0^2/n} \cdot \bar{x} + \frac{1}{\Delta_0^2} \cdot \theta_0}{\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2}} = 0{,}691,$$

$$\tilde{\Delta}_0^2 = \mathbf{D}(\theta | x_1, ..., x_n) = \left(\frac{1}{\sigma_0^2/n} + \frac{1}{\Delta_0^2} \right)^{-1} = 0{,}015.$$

Соответственно:

$$\hat{\theta}^{(\text{Б})} = \mathbf{E}(\theta | x_1, ..., x_n) = 0{,}691$$

и с вероятностью $P_0 = 0{,}95$ можем утверждать, что $\hat{\theta}^{(\text{Б})} - u_{0,025} \cdot \tilde{\Delta}_0 < \theta < \hat{\theta}^{(\text{Б})} + u_{0,025} \cdot \tilde{\Delta}$.

С учетом того, что 2,5%-ная точка стандартного нормального распределения $u_{0,025} = 1{,}96$ и $\tilde{\Delta}_0 = \sqrt{\tilde{\Delta}_0^2} = 0{,}120$, имеем:

$$\theta \in [0{,}451; 0{,}931] \text{ с вероятностью } P_0 = 0{,}95.$$

Решение этих же задач, основанное на *методе максимального правдоподобия*, дает:

$$\hat{\theta}_{\text{мп}} = \bar{x} = 0{,}775 \text{ и } \theta \in [0{,}447; 1{,}103] \text{ с вероятностью } P_0 = 0{,}95$$

(концы последнего доверительного интервала вычислены по формулам $\hat{\theta}_{\text{мп}} \pm u_{0,025} \cdot \frac{\sigma_0}{\sqrt{n}}$).

Мы видим, что использование априорной информации о неизвестном значении параметра $\theta = \mathbf{E}(\ln \eta)$ и применение, соответственно, байесовского подхода в данной задаче позволили уточнить оценку и, в частности, сузить интервальную оценку по сравнению с классическим подходом почти в полтора раза.

Задача 2. Оценка интенсивности вызовов, поступающих на пункт «Скорой помощи» в час.

Число вызовов ξ, поступающих на пункт «Скорой помощи» в час, описывается распределением Пуассона с неизвестным значением параметра $\theta = \mathbf{E}\xi$ (см. выше, пример 5). Результаты регистрации числа вызовов x_i (в час), зафиксированные в течение одной смены (длящейся 8 часов), приведены в следующей таблице:

i	1	2	3	4	5	6	7	8
x_i	3	1	4	2	6	3	3	2

Из опыта работы аналогичных пунктов определено априорное среднее значение $\theta_0 = \mathbf{E}\theta = 3{,}6$, причем случайный разброс значений этого параметра характеризуется дисперсией $\Delta_0^2 = \mathbf{D}\theta = 0{,}09$.

Требуется:

Используя сопряженное априорное распределение параметра θ, получить байесовские точечную и интервальную (с уровнем доверия $P_0 = 0{,}95$) оценки средней интенсивности $\theta = \mathbf{E}\xi$ вызовов, поступающих на пункт «Скорой помощи», и сравнить их с соответствующими оценками метода максимального правдоподобия.

Решение. Как было установлено выше (см. пример 5), сопряженное априорное распределение параметра θ в этом случае существует и описывается гамма-законом, параметры α и β которого определяются (в соответствии с рекомендациями п. 3.3) из системы уравнений:

$$\begin{cases} \mathbf{E}\theta = \dfrac{\alpha}{\beta} = 3{,}60 \\ \mathbf{D}\theta = \dfrac{\alpha}{\beta^2} = 0{,}09 \end{cases}.$$

Отсюда $\alpha = 144$ и $\beta = 40$.

В соответствии с выведенными выше (см. п.4) формулами пересчета (43) параметров апостериорного гамма-распределения имеем:

$$\tilde{\alpha} = \alpha + \sum_{i=1}^{8} x_i = 144 + 24 = 168,$$

$$\tilde{\beta} = \beta + n = 40 + 8 = 48.$$

Таким образом:

$$\hat{\theta}^{(Б)} = E(\theta | x_1, \ldots, x_n) = \frac{\tilde{\alpha}}{\tilde{\beta}} = 3{,}5,$$

и можно утверждать, что с вероятностью $P_0 = 0{,}95$ справедливы неравенства

$$\gamma_{0,975}(\tilde{\alpha}, \tilde{\beta}) < \theta < \gamma_{0,025}(\tilde{\alpha}, \tilde{\beta}),$$

где $\gamma_q(\tilde{\alpha}, \tilde{\beta})$ — $100q$%-ная точка гамма-распределения с параметрами $\tilde{\alpha}$ и $\tilde{\beta}$. Воспользовавшись известными формулами (см. приложение 2):

$$\gamma_q(\tilde{\alpha}; \tilde{\beta}) = \frac{1}{2\tilde{\beta}} \chi_q^2(2\tilde{\alpha})$$

и, при $m > 100$:

$$\chi_q^2(m) \approx m + u_q \cdot \sqrt{2m},$$

где $\chi_q^2(m)$ и u_q — $100q$%-ные точки «хи-квадрат»-распределения и стандартного нормального распределения, соответственно, имеем:

$$\theta \in [2{,}97;\ 4{,}03] \text{ с вероятностью } P_0 = 0{,}95.$$

Решение этих же задач, основанное на методе максимального правдоподобия, дает:

$$\hat{\theta}_{мп} = \bar{x} = 3{,}0;$$

$$\theta \in [1{,}80;\ 4{,}20;]\ \text{ с вероятностью }\ P_0 = 0{,}95^3.$$

Мы видим, что в данном случае использование априорной информации о параметре θ в рамках байесовского подхода позволило сузить размах интервальной оценки более чем в два раза!

Задача 3. Оценка «необходимой доли брака» θ в продукции, производимой автоматической линией.

Предприятие приобрело новую автоматическую линию (АЛ). Для оценки так называемой «необходимой доли брака» θ — вероятности того, что произведенное этой АЛ в режиме стационарного функционирования изделие окажется некондиционным, — было проконтролировано $n = 5$ партий по $M = 80$ изделий в каждой партии. Число дефектных изделий ξ, обнаруженных в партии изделий объема M, адекватно описывается биномиальным з.р.в. с параметрами θ и M. Результаты контроля представлены в таблице (x_i — это число дефектных изделий, обнаруженных в i-й проконтролированной партии):

i	1	2	3	4	5
x_i	2	0	3	1	2

Кроме того, проведенный анализ работы аналогичных АЛ, установленных на других предприятиях, показал, что «необходимая доля брака» в среднем равна 0,01 и имеет разброс, характеризуемый среднеквадратическим отклонением 0,003.

Требуется:

Используя сопряженное априорное распределение параметра θ, получить байесовские точечную и интервальную (с уровнем доверия $P_0 = 0{,}90$) оценки «необходимой доли брака» θ и сравнить их с соответствующими оценками метода максимального правдоподобия.

Решение. Выше (см. п. 3) было установлено, что сопряженное априорное распределение параметра θ в данном случае существует и описывается бета-распределением, параметры a и b которого определяются из системы (см. п.3.3):

$$\begin{cases} \dfrac{a}{a+b} = 0{,}01 \\ \dfrac{ab}{(a+b)^2(a+b+1)} = (0{,}003)^2 \end{cases}.$$

Решение этой системы дает $a = 10$ и $b = 990$. Воспользовавшись формулами пересчета (41), получаем значения параметров $\tilde{a}$ и $\tilde{b}$ *апостериорного* распределения θ:

$$\tilde{a} = a + \sum_{i=1}^{5} x_i = 10 + 8 = 18,$$

$$\tilde{b} = b + 5 \cdot 80 - \sum_{i=1}^{5} x_i = 1382.$$

[3] Данная интервальная оценка основана на асимптотической $\left(\theta; \sqrt{\frac{\theta}{n}}\right)$-нормальности оценки максимального правдоподобия $\hat{\theta}_{\text{мп}}$.

Таким образом:

$$\hat{\theta}^{(Б)} = \mathbf{E}(\theta | x_1, \ldots, x_n) = \frac{\tilde{a}}{\tilde{a} + \tilde{b}} = 0{,}01286,$$

и можно утверждать, что с вероятностью $P_0 = 0{,}90$ справедливы неравенства

$$\beta_{0,95}(\tilde{a}, \tilde{b}) < \theta < \beta_{0,05}(\tilde{a}, \tilde{b}),$$

где $\beta_q(\tilde{a}, \tilde{b})$ — $100q$%-ная точка бета-распределения с параметрами $\tilde{a}$ и $\tilde{b}$. Воспользовавшись известными равенствами

$$\beta_q(\tilde{a}; \tilde{b}) = \frac{\tilde{a}\, F_q(2\tilde{a}; 2\tilde{b})}{\tilde{b} + \tilde{a}\, F_q(2\tilde{a}; 2\tilde{b})}, \quad F_{1-q}(\nu_1, \nu_2) = \frac{1}{F_q(\nu_2, \nu_1)}$$

и таблицами $100q$-процентных точек $F_q(\nu_1, \nu_2)$ распределения F с числами степеней свободы числителя ν_1 и знаменателя ν_2, имеем:

$$\theta \in [0{,}0083; 0{,}0182] \text{ с вероятностью } P_0 = 0{,}90.$$

Решение этих же задач, основанное на методе максимального правдоподобия (см., например, [Айвазян, Мхитарян (2001б), задача 1.18]), дает:

$$\hat{\theta}_{\text{мп}} = \frac{1}{nM} \sum_{i=1}^{n} x_i = 0{,}02,$$

$$\theta \in [0{,}0085; 0{,}0315] \text{ с вероятностью } P_0 = 0{,}90.$$

Размах этой интервальной оценки в 2,3 раза превосходит ширину *байесовской* интервальной оценки!

Задача 4. Оценка интервала движения автобуса.

Приходящий в случайные моменты времени на остановку пассажир в течение пяти своих поездок фиксировал время ожидания автобуса (в минутах): $x_1 = 1{,}2$; $x_2 = 2{,}5$; $x_3 = 0{,}5$; $x_4 = 3{,}2$; $x_5 = 2{,}9$. Известно, что автобус ходит строго по расписанию с интервалом в θ минут, так что время ожидания автобуса пассажиром можно считать случайной величиной ξ, подчиненной $[0; \theta]$-равномерному з.р.в. (см. выше, пример 7). Пытаясь оценить интервал движения автобуса, пассажир сумел получить дополнительную информацию о параметре θ: из анализа опыта работы различных автобусных маршрутов города, функционирующих в едином регламентном режиме, следовало, что среднее значение этого параметра равно 5,38 мин., а случайный разброс в его значениях характеризуется средним квадратическим отклонением, равным 1,39 мин.

Требуется:

Используя сопряженное априорное распределение параметра θ, получить байесовские точечную и интервальную (с уровнем доверия $P_0 = 0{,}95$) оценки для неизвестного интервала движения автобуса и сравнить их с соответствующими оценками, основанными на методе максимального правдоподобия.

Решение. В п. 3 (см. пример 7) было установлено, что сопряженное априорное распределение параметра θ в данном случае существует и описывается распределением Парето с параметром формы α и параметром сдвига $\theta_{\min}$, которые определяются из системы уравнений (32), т. е. по формулам (32'). В нашем случае имеем:

$$\alpha = 1+\sqrt{1+\frac{\theta_0^2}{\Delta^2}} = 1+\sqrt{1+\frac{5{,}38^2}{1{,}39^2}} = 5{,}00,$$

$$\theta_{\min} = \frac{1}{\alpha}\theta_0 \cdot (\alpha-1) = \frac{1}{5}\cdot 5{,}38 \cdot 4 = 4{,}30 \text{ (мин.)}.$$

Параметры *апостериорного* распределения Парето определяются формулами пересчета (45):

$$\tilde{\alpha} = \alpha + n = 5+5 = 10,$$

$$\tilde{\theta}_{\min} = \max\{\theta_{\min}; x_1, \ldots, x_5\} = 4{,}30 \text{ (мин.)}.$$

Соответственно:

$$\hat{\theta}^{(Б)} = \mathbf{E}(\theta | x_1, \ldots, x_5) = \frac{\tilde{\alpha}\cdot\tilde{\theta}_{\min}}{\tilde{\alpha}-1} = 4{,}78 \text{ (мин.)}$$

и $\theta \in [\theta_{0,975}(\tilde{\alpha}; \tilde{\theta}_{\min}); \theta_{0,025}(\tilde{\alpha}; \tilde{\theta}_{\min})]$ с вероятностью $P_0 = 0{,}95$, где $\theta_q(\tilde{\alpha}; \tilde{\theta}_{\min})$ — это $100q$%-ная точка распределения Парето с параметрами $(\tilde{\alpha}; \tilde{\theta}_{\min})$. Поскольку функция распределения Парето определяется соотношением

$$F(\theta) = P\{\xi(\tilde{\alpha}; \tilde{\theta}_{\min}) < \theta\} = 1-\left(\frac{\tilde{\theta}_{\min}}{\theta}\right)^{\tilde{\alpha}},$$

то значения $\theta_{0,975}(\tilde{\alpha}; \tilde{\theta}_{\min})$ и $\theta_{0,025}(\tilde{\alpha}; \tilde{\theta}_{\min})$ определяется из уравнений соответственно:

$$\left(\frac{\tilde{\theta}_{\min}}{\theta_{0,975}(\tilde{\alpha}; \tilde{\theta}_{\min})}\right)^{\tilde{\alpha}} = 0{,}975,$$

$$\left(\frac{\tilde{\theta}_{\min}}{\theta_{0,025}(\tilde{\alpha}; \tilde{\theta}_{\min})}\right)^{\tilde{\alpha}} = 0{,}025.$$

Решение этих уравнений относительно $\theta_{0,975}(\tilde{\alpha}; \tilde{\theta}_{\min})$ и $\theta_{0,025}(\tilde{\alpha}; \tilde{\theta}_{\min})$ при $\tilde{\alpha} = 10$ и $\tilde{\theta}_{\min} = 4{,}3$ дает:

$$\theta_{0,975}(\tilde{\alpha}; \tilde{\theta}_{\min}) = 4{,}31 \text{ и } \theta_{0,025}(\tilde{\alpha}; \tilde{\theta}_{\min}) = 6{,}22,$$

так что

$$\theta \in [4{,}31; 6{,}22] \text{ с вероятностью } P_0 = 0{,}95.$$

Решение тех же задач, основанное на *методе максимального правдоподобия,* дает (см. [Айвазян, Мхитарян (2001б), задача 1.22]):

$$\hat{\theta}_{\text{мп}} = 3{,}84 \text{ (мин.)},$$

$$\theta \in [3{,}22; 6{,}69] \text{ с вероятностью } P_0 = 0{,}95$$

(здесь дается оценка максимального правдоподобия $\hat{\theta}_{\text{мп}}$, *подправленная на несмещенность*). И в данном случае байесовский подход позволил сузить ширину доверительного интервала почти в 2 раза (точнее, в 1,82 раза).

Задача 5. Оценка параметров модели зависимости душевых доходов от объема автономных инвестиций.

В нижеследующей таблице приведены макроэкономические данные по США, характеризующие среднедушевой доход y_t и автономные инвестиции x_t (в долларах, в дефлированных, с помощью индекса стоимости жизни, ценах) за 1922–1941 годы. Инвестиции определены приближенно как разность между среднедушевым доходом и душевым расходом на личное потребление (данные заимствованы из работы *Haavelmo T.* Methods of Measuring the Marginal Propensity to Consume. — JASA, vol. 42 (1947), pp 105–122).

t	1 (1922)	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20 (1941)
x_t	39	60	42	52	47	51	45	60	39	41	22	17	27	33	48	51	33	46	54	100
y_t	433	483	479	486	494	498	511	534	478	440	372	381	419	449	511	520	477	517	548	629

Анализируется нормальная классическая линейная модель парной регрессии (см. пример 9 в пункте 2 при $k = 1$):

$$y_t = \theta_0 + \theta_1 x_t + \varepsilon_t,\ \ t = 1, 2, \ldots, 20. \qquad (48)$$

Анализ предыстории и экспертных оценок модели позволил получить следующую априорную информацию о значениях параметров θ_0, θ_1 и $h = (\mathbf{D}\varepsilon_t)^{-1}$:

$$\theta_0^0 = \mathbf{E}\theta_0 = 330; \quad \theta_1^0 = \mathbf{E}\theta_1 = 2{,}85; \quad h_0 = \mathbf{E}h = 0{,}002;$$

$$\Delta_0^2 = \mathbf{D}\theta_0 = 225; \quad \Delta_1^2 = \mathbf{D}\theta_1 = 0{,}01; \quad \Delta_h^2 = \mathbf{D}h = 25 \cdot 10^{-8}.$$

Требуется:

Используя сопряженное априорное распределение параметров $(\theta_0; \theta_1; h)$, получить байесовские точечные и интервальные (с уровнем доверия $P_0 = 0{,}90$) оценки этих параметров и сравнить их с соответствующими оценками метода максимального правдоподобия.

Решение. Проведенный в пунктах 2 и 3 анализ примера 9 показал, что в данном случае существует сопряженное с наблюдаемой генеральной совокупностью распределение параметров $(\theta_0; \theta_1; h)$ и что оно описывается трехмерным гамма-нормальным распределением (29) с параметрами $\Theta^0 = (\theta_0^0; \theta_1^0)^{\mathrm{T}}$, Λ_0, α и β, определяемыми в соответствии с рекомендациями (33) и (36), т. е.: $\Theta^0 = (330; 2{,}85)^{\mathrm{T}}$; $\alpha = \dfrac{h_0^2}{\Delta_h^2} = 16$; $\beta = \dfrac{h_0}{\Delta_h^2} = 8000$;

$$\Lambda_0 = \begin{pmatrix} 2{,}37 & 0 \\ 0 & 53333{,}3 \end{pmatrix}.$$

Параметры *апостериорного* гамма-нормального распределения вычисляются в соответствии с формулами пересчета (47):

$$\tilde{\Theta}_0 = \begin{pmatrix} 349{,}0 \\ 2{,}9 \end{pmatrix},\ \tilde{\alpha} = 26,\ \tilde{\beta} = 14578,\ \tilde{\Lambda}_0 = \begin{pmatrix} 22{,}37 & 907 \\ 907 & 100\,176 \end{pmatrix}.$$

Точечные байесовские оценки параметров $(\theta_0; \theta_1; h)$ определяются средними значениями соответствующих частных апостериорных распределений. С учетом свойств (*i*) и (*ii*) многомерного гамма-нормального распределения (см. выше, п. 5 раздела 3.3) имеем:

$$\hat{\Theta}^{(Б)} = \mathbf{E}(\Theta|\mathbf{X},Y) = \widetilde{\Theta}_0 = (349{,}0;\, 2{,}90)^{\mathrm{T}},$$

$$\hat{h}^{(Б)} = \mathbf{E}(h|\mathbf{X},Y) = \frac{\widetilde{\alpha}}{\widetilde{\beta}} = 0{,}00178.$$

При выводе *интервальных* байесовских оценок также используются свойства (*i*) и (*ii*) многомерного гамма-нормального распределения, а также факт $t(2\widetilde{\alpha})$-распределенности случайных величин $(\hat{\theta}_j^{(Б)} - \theta_j)\sqrt{\widetilde{c}_j}$, где параметр точности $\widetilde{c}_j$ вычисляется по блочным компонентам матрицы точности $\widetilde{B}$ частного апостериорного обобщенного многомерного $t(2\widetilde{\alpha}|\hat{\Theta}^{(Б)};\widetilde{B})$-распределения по формуле

$$\widetilde{c}_j = \widetilde{b}_{jj} - \widetilde{B}_{j.} \cdot \widetilde{B}(j) \cdot \widetilde{B}_{.j} \tag{49}$$

(см. Приложение 1б). Участвующие в этом соотношении число $\widetilde{b}_{jj}$, $1\times(k-1)$-матрица $\widetilde{B}_{j.}$, $(k-1)\times 1$-матрица $\widetilde{B}_{.j}$ и $(k-1)\times(k-1)$-матрица $\widetilde{B}(j)$ определяются следующим блочным представлением матрицы $\widetilde{B}$:

$$\widetilde{B} = \begin{pmatrix} \widetilde{b}_{jj} & \widetilde{B}_{j.} \\ \widetilde{B}_{.j} & \widetilde{B}(j) \end{pmatrix}. \tag{50}$$

В нашем случае $k = 2$, $j = 0$ или 1, матрица $\widetilde{B} = \dfrac{\widetilde{\alpha}}{\widetilde{\beta}}\widetilde{\Lambda}_0 = \begin{pmatrix} 0{,}040 & 1{,}618 \\ 1{,}618 & 178{,}665 \end{pmatrix}$, так что $\widetilde{c}_0 = 0{,}040 - (1{,}618)^2/178{,}665 = 0{,}0254$ и $\widetilde{c}_1 = 178{,}665 - (1{,}618)^2/0{,}040 = 113{,}217$. Следовательно, с вероятностью $P_0 = 0{,}90$ мы можем утверждать, что $|\hat{\theta}_0^{(Б)} - \theta_0|\cdot\sqrt{0{,}0254} < t_{0{,}05}(52)$ и $|\hat{\theta}_1^{(Б)} - \theta_1|\cdot\sqrt{113{,}217} < t_{0{,}05}(52)$, так что (с учетом того, что $t_{0{,}05}(52) = 1{,}676$) имеем:

$$\theta_0 \in [338{,}5;\, 359{,}5] \text{ с вероятностью } P_0 = 0{,}90,$$

$$\theta_1 \in [2{,}743;\, 3{,}057] \text{ с вероятностью } P_0 = 0{,}90.$$

Поскольку параметр h подчиняется апостериорному гамма-распределению с параметрами $\widetilde{\alpha}$ и $\widetilde{\beta}$, то

$$h \in \left[\gamma_{0{,}95}(\widetilde{\alpha};\widetilde{\beta});\, \gamma_{0{,}05}(\widetilde{\alpha};\widetilde{\beta})\right] \text{ с вероятностью } P_0 = 0{,}90.$$

Используя соотношение $\gamma_q(\widetilde{\alpha};\widetilde{\beta}) = \dfrac{1}{2\widetilde{\beta}}\chi_q^2(2\widetilde{\alpha})$, имеем (с учетом $\chi^2_{0{,}95}(52) \approx 36{,}4$ и $\chi^2_{0{,}05}(52) \approx$ $\approx 69{,}8$):

$$h \in [0{,}00125;\, 0{,}00239] \text{ с вероятностью } P_0 = 0{,}90.$$

Оценивание модели (48) с помощью *метода максимального правдоподобия* (дающего в данном случае те же результаты, что и *метод наименьших квадратов*) приводит к следующим точечным и интервальным оценкам:

$$\hat{\Theta}_{\text{мп}} = \hat{\Theta}_{\text{мнк}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}Y = (344{,}7;\, 3{,}05)^{\mathrm{T}},$$

$$\hat{h}_{\text{мп}} = \left[\frac{1}{18}(Y - \mathbf{X}\hat{\Theta}_{\text{мп}})^{\mathrm{T}}(Y - \mathbf{X}\hat{\Theta}_{\text{мп}})\right]^{-1} = 0{,}0015,$$

$$316{,}1 < \theta_0 < 373{,}3;\ 2{,}45 < \theta_1 < 3{,}64 \text{ с вероятностью } P_0 = 0{,}90,$$

$$h \in [0{,}00093;\, 0{,}00285] \text{ с вероятностью } P_0 = 0{,}90.$$

Мы видим, что байесовский подход позволяет сузить доверительный интервал для θ_0 в 2,6 раза, для θ_1 — в 3,7 раза и для h — в 1,4 раза по сравнению с подходом, основанным на методе максимального правдоподобия.

6. Байесовский прогноз зависимой переменной, основанный на нормальной классической линейной модели множественной регрессии

Мы продолжаем рассматривать нормальную КЛММР

$$y_t = \theta_0 + \sum_{j=1}^{k} \theta_j \cdot x_t^{(j)} + \varepsilon_t, \ t = 1, 2, \ldots, n,$$

или, в матричной записи, модель (8)–(9) (см. выше), в которой остатки $\varepsilon_i = \varepsilon(X_i)$ нормальны, гомоскедастичны и взаимнонекоррелированы при **любом** (*а не только наблюденном*) наборе значений объясняющих переменных.

Введем в рассмотрение, наряду с наблюденными значениями $\mathbf{X}$ и Y анализируемых переменных $X = (1; x^{(1)}, x^{(2)}, \ldots, x^{(k)})^{\mathrm{T}}$ и y, их прогнозные (на q тактов времени вперед) значения:

$$\widetilde{\mathbf{X}} = \begin{pmatrix} 1 & x_{n+1}^{(1)} & x_{n+1}^{(2)} & \ldots & x_{n+1}^{(k)} \\ 1 & x_{n+2}^{(1)} & x_{n+2}^{(2)} & \ldots & x_{n+2}^{(k)} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n+q}^{(1)} & x_{n+q}^{(2)} & \ldots & x_{n+q}^{(k)} \end{pmatrix} \text{ и } \widetilde{Y} = \begin{pmatrix} y_{n+1} \\ y_{n+2} \\ \vdots \\ y_{n+q} \end{pmatrix},$$

а также соответствующие остатки $\widetilde{\varepsilon} = (\varepsilon_{n+1}, \varepsilon_{n+2}, \ldots, \varepsilon_{n+q})^{\mathrm{T}}$.

Тогда в соответствии с (8)–(9):

$$\begin{cases} \widetilde{Y} = \widetilde{\mathbf{X}}\Theta + \widetilde{\varepsilon} \\ \widetilde{\varepsilon} \in N_q(\mathbf{0}; h^{-1} \cdot \mathbf{I}_q) \end{cases}.$$

Для того чтобы строить точечные и интервальные оценки для $\widetilde{Y}$ по заданным значениям $\mathbf{X}, \widetilde{\mathbf{X}}$ и Y, очевидно, надо располагать плотностью условного распределения $p(\widetilde{Y} | \mathbf{X}; \widetilde{\mathbf{X}}; Y)$, которую обычно называют «**прогнозной функцией плотности вероятности**». Но поскольку из (8а)–(9а) следует, что распределение вектора $\widetilde{Y}$ зависит также от параметров Θ и h, а *они в байесовском подходе интерпретируются как случайные величины, имеющие соответствующее апостериорное распределение*, то реализуется следующая схема определения прогнозной функции плотности $p(\widetilde{Y} | \mathbf{X}; \widetilde{\mathbf{X}}; Y)$:

$$p(\widetilde{Y}|\mathbf{X};\widetilde{\mathbf{X}};Y) = \int_{\Theta}\int_{h} p(\widetilde{Y};\Theta;h|\mathbf{X};\widetilde{\mathbf{X}};Y)d\Theta dh = \int_{\Theta}\int_{h} p(\widetilde{Y}|\Theta;h;\mathbf{X};\widetilde{\mathbf{X}};Y) \cdot p(\Theta;h|\mathbf{X};\widetilde{\mathbf{X}};Y)\, d\Theta dh. \quad (51)$$

Правая часть (51) получена с использованием формулы произведения условных вероятностей $P(AB|C) = P(A|B,C) \cdot P(B|C)$. С учетом того, что $p(\widetilde{Y}|\Theta;h;\mathbf{X};\widetilde{\mathbf{X}};Y) = p(\widetilde{Y}|\Theta;h;\widetilde{\mathbf{X}}) \sim h^{\frac{q}{2}} \cdot e^{-\frac{h}{2}(\widetilde{Y}-\mathbf{X}\Theta)^{\mathrm{T}}(\widetilde{Y}-\mathbf{X}\Theta)}$, а $p(\Theta;h|\mathbf{X};\widetilde{\mathbf{X}};Y) = p(\Theta;h|\mathbf{X};Y)$ — гамма-нормальное распределение с параметрами $\widetilde{\Theta}_0$, $\widetilde{\Lambda}_0$, $\widetilde{\alpha}$ и $\widetilde{\beta}$, определяемыми по параметрам $\Theta_0, \Lambda_0, \alpha$ и β априорного гамма-нормального распределения $p(\Theta, h)$ по формулам (47), интегрирование в правой части (51) дает:

$$p(\widetilde{Y}|\mathbf{X};\widetilde{\mathbf{X}};Y) \sim \left[1+\frac{1}{v}(\widetilde{Y}-\widetilde{\mathbf{X}}\widetilde{\Theta}_0)^{\mathrm{T}}B'(Y-\widetilde{\mathbf{X}}\widetilde{\Theta}_0)\right]^{-\frac{v+q}{2}}, \tag{52}$$

где $v = n-k-1$ и $B' = \dfrac{\widetilde{\alpha}}{\widetilde{\beta}}\left[\mathbf{I}_q - \widetilde{\mathbf{X}}(\Lambda_0 + \mathbf{X}^{\mathrm{T}}\mathbf{X} + \widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\right]$ (53)

(подробное доказательство этого факта читатель найдет, например, в [Зельнер (1980)]). Таким образом, мы пришли к тому, что условное распределение q-мерного вектора $\widetilde{Y}$ при заданных значениях $\mathbf{X}, Y$ и $\widetilde{\mathbf{X}}$ описывается обобщенным многомерным t-распределением с $n-k-1$степенями свободы, параметром сдвига $\widetilde{\mathbf{X}}\widetilde{\Theta}_0$ и матрицей точности B', определенной соотношением (53) (т. е. $(\widetilde{Y}|\mathbf{X};\widetilde{\mathbf{X}};Y) = t(n-k-1\|\widetilde{\mathbf{X}}\widetilde{\Theta}_0;B')$, см. Приложение 1б).

Используя известные свойства обобщенного t-распределения Стьюдента (см. Приложение 1б), получаем следующие байесовские прогнозы для $\widetilde{Y}$:

- *точечный байесовский прогноз для компонент вектора* $\widetilde{Y}$ определяется соотношением:

$$\hat{y}_{n+m}(\text{прогнозное}) = (\hat{\Theta}^{(\text{Б})})^{\mathrm{T}} \cdot X_{n+m},\ m = 1,2,\ldots,q; \tag{54}$$

- *интервальный байесовский прогноз для компонент вектора* $\widetilde{Y}$ *с вероятностью* P_0 определяется соотношением:

$$y_{n+m} \in \left[\hat{y}_{n+m} - t_{\frac{1-P_0}{2}}(n-k-1)\cdot\frac{1}{\sqrt{c'_m}};\quad \hat{y}_{n+m} + t_{\frac{1-P_0}{2}}(n-k-1)\cdot\frac{1}{\sqrt{c'_m}}\right],\ m = 1, 2, \ldots, q, \tag{55}$$

где $t_\varepsilon(v)$ — $100\varepsilon\%$-ная точка стандартного $t(v)$-распределения Стьюдента, а величины c'_m вычисляются по схеме (49)—(50) с заменой $(k\times k)$-матрицы $\widetilde{B}$ на $q\times q$-матрицу B', определенную соотношением (53);

- *байесовская прогнозная доверительная область* $\Delta\widetilde{Y}$ *для вектора* $\widetilde{Y} = (y_{n+1},\ldots,y_{n+q})^{\mathrm{T}}$ состоит, с заданной вероятностью P_0, из всех тех $\widetilde{Y} = (y_{n+1},\ldots,y_{n+q})^{\mathrm{T}}$, которые удовлетворяют неравенству

$$\frac{1}{q}(\widetilde{Y}-\widetilde{\mathbf{X}}\hat{\Theta}^{(\text{Б})})^{\mathrm{T}}\Sigma_{\widetilde{Y}}^{-1}(\widetilde{Y}-\widetilde{\mathbf{X}}\hat{\Theta}^{(\text{Б})}) < F_{1-P_0}(q;\ n-k-1), \tag{56}$$

где $F_\varepsilon(v_1,v_2)$ — $100\varepsilon\%$-ная точка $F(v_1,v_2)$-распределения, $\hat{\Theta}^{(\text{Б})}$ — байесовская точечная оценка параметров регрессии Θ, а $\Sigma_{\widetilde{Y}} = \dfrac{n-k-1}{n-k-3}(B')^{-1}$ — ковариационная матрица вектора $\widetilde{Y}$. Можно показать, что для модели (8а)–(9а) эта область имеет форму q-мерного эллипсоида.

Рассмотрим реализацию описанной выше схемы построения точечных и интервальных байесовских прогнозов значений зависимой переменной в нормальной КЛММР на нашем примере, проанализированном в задаче 5.

Задача 5 (*продолжение*).

В условиях примера, рассмотренного выше в задаче 5, *требуется*:

по заданным (планируемым) значениям автономных инвестиций $x_{21} = 120$ и $x_{22} = 140$ построить точечные и интервальные байесовские прогнозы для среднедушевых доходов населения y_{21} и y_{22}, а также прогнозную доверительную область $\Delta\widetilde{Y}$ для этих значений с уровнем доверия $P_0 = 0{,}90$. Сравнить полученные решения с решениями, основанными на методе максимального правдоподобия.

Решение. Итак, в нашем случае:

$$\tilde{\mathbf{X}} = \begin{pmatrix} 1 & 120 \\ 1 & 140 \end{pmatrix}; \quad \tilde{Y} = \begin{pmatrix} y_{21} \\ y_{22} \end{pmatrix} = ?$$

В соответствии с (52) плотность условного распределения вектора $\tilde{Y}$ при заданных $\mathbf{X}, \tilde{\mathbf{X}}$ и Y описывается обобщенным многомерным t-распределением с числом степеней свободы $v = 20 - 1 - 1 = 18$, параметром сдвига $\begin{pmatrix} 1 & 120 \\ 1 & 140 \end{pmatrix}\begin{pmatrix} 349{,}0 \\ 2{,}9 \end{pmatrix}$ и матрицей точности B', определенной соотношением (53).

Произведя необходимые вычисления по формулам (53)–(56) и используя известные свойства обобщенного многомерного t-распределения (см. Приложение 1б), имеем:

$$\hat{y}_{21}(\text{прогн.}) = 349 + 2{,}9 \cdot 120 = 697{,}4,$$

$$\hat{y}_{22}(\text{прогн.}) = 349 + 2{,}9 \cdot 140 = 755{,}5,$$

$$B' = \begin{pmatrix} 0{,}00159 & -0{,}00022 \\ -0{,}00022 & 0{,}00152 \end{pmatrix}; \quad \Sigma_{\tilde{Y}} = \begin{pmatrix} 721{,}8 & 106{,}8 \\ 106{,}8 & 757{,}4 \end{pmatrix},$$

$$y_{21} \in [653{,}6;\ 741{,}2] \text{ с вероятностью } P_0 = 0{,}90,$$

$$y_{22} \in [710{,}7;\ 800{,}3] \text{ с вероятностью } P_0 = 0{,}90,$$

$$\Delta_{\tilde{Y}} = \left\{ \begin{pmatrix} y_{21} \\ y_{22} \end{pmatrix} : \frac{1}{2} \begin{pmatrix} y_{21} - 697{,}4 \\ y_{22} - 755{,}5 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 0{,}00159 & -0{,}00022 \\ -0{,}00022 & 0{,}00152 \end{pmatrix} \begin{pmatrix} y_{21} - 697{,}4 \\ y_{22} - 755{,}5 \end{pmatrix} < 2{,}62 \right\}.$$

Сравним эти результаты с соответствующими прогнозами, основанными на оценках *метода максимального правдоподобия*:

- точечный прогноз

$$\hat{y}_{21}^{\text{мп}}(\text{прогнозное}) = 344{,}7 + 3{,}05 \cdot 120 = 710{,}7,$$

$$\hat{y}_{22}^{\text{мп}}(\text{прогнозное}) = 344{,}7 + 3{,}05 \cdot 140 = 771{,}7;$$

- интервальный прогноз строится на основе $t(n-2)$-распределенности случайных величин

$$\frac{\hat{y}_{n+m}^{\text{мп}}(\text{прогн.}) - y_{n+m}}{\hat{\sigma}_{\text{мп}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+m} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}}, \quad m = 1, 2;$$

в нашем случае $n = 20$, $x_{n+1} = 120$, $x_{n+2} = 140$, $\hat{\sigma}^2_{\text{мп}} = \frac{1}{\hat{h}_{\text{мп}}} = \frac{1}{0{,}0015} = 666{,}67$; $\hat{\sigma}_{\text{мп}} = 25{,}82$; $\frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{20}(x_i - \bar{x})^2} = \frac{5572{,}6}{5711} = 0{,}976$; $\frac{(x_{n+2} - \bar{x})^2}{\sum_{i=1}^{20}(x_i - \bar{x})^2} = \frac{8958{,}6}{5711} = 1{,}569$ и $t_{0,05}(18) = 1{,}734$, так что:

$$y_{21} \in [647{,}1;\ 774{,}3] \text{ с вероятностью } P_0 = 0{,}90,$$

$$y_{22} \in [699{,}2;\ 844{,}2] \text{ с вероятностью } P_0 = 0{,}90.$$

Мы видим, что и в прогнозе байесовский подход позволяет сузить ширину прогнозной интервальной оценки для y_{22} в 1,45 раза, а для y_{22} — в 1,62 раза!

Приложения

Некоторые сведения об одномерных и многомерных законах распределения вероятностей, используемые в байесовском подходе

Приложение 1а

Обобщенное одномерное распределение Стьюдента с ν степенями свободы, параметром сдвига θ_0 и параметром точности c ($t(\nu|\theta_0; c)$-распределение)

Как известно (см., например, [Айвазян, Мхитарян (2001), гл. 3]), *стандартный* з.р.в. Стьюдента с ν степенями свободы (ст. св.) описывает распределение случайной величины

$$t(\nu) = \frac{\xi_0}{\sqrt{\frac{1}{\nu}\sum_{i=1}^{\nu}\xi_1^2}}, \tag{П.1}$$

где $\xi_0, \xi_1, \ldots, \xi_\nu$ — статистически взаимонезависимые $(0; \sigma^2)$-нормально распределенные случайные величины. Значение соответствующей функции плотности вероятности $f_{t(\nu)}(x)$ в точке x задается соотношением:

$$f_{t(\nu)}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\cdot\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{П.2}$$

причем $\mathbf{E}t(\nu) = 0$ и $\mathbf{D}t(\nu) = \dfrac{\nu}{\nu-2}$ $(\nu > 2)$.

Введем в рассмотрение случайную величину $t(\nu|\theta_0; c)$, являющуюся линейной функцией от $t(\nu)$, а именно:

$$t(\nu|\theta_0; c) = \frac{1}{\sqrt{c}}t(\nu) + \theta_0. \tag{П.3}$$

Легко показать, что функция плотности вероятности $f_{t(\nu|\theta_0;c)}(x)$ случайной величины $t(\nu|\theta_0; c)$ в точке x имеет вид:

$$f_{t(\nu|\theta_0;c)}(x) = \frac{\sqrt{c}\cdot\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\cdot\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{c(x-\theta_0)^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \tag{П.4}$$

причем $\mathbf{E}t(\nu|\theta_0; c) = \theta_0$ и $\mathbf{D}t(\nu|\theta_0; c) = \dfrac{1}{c}\cdot\dfrac{\nu}{\nu-2}$ $(\nu > 2)$.

Распределение, задаваемое плотностью (П.4), называют *обобщенным распределением Стьюдента* (*или* $t(\nu|\theta_0\,;c)$-*распределением*) с параметром сдвига θ_0 и параметром *точности* c. Отметим, что в данном случае параметр точности c не есть величина, обратная к дисперсии случайной величины $t(\nu|\theta_0\,;c)$: дисперсия может и не существовать (при $\nu \le 2$).

Приложение 1б

Обобщенное k-мерное ($k \ge 2$) распределение Стьюдента с ν степенями свободы, параметром сдвига $\Theta_0 = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)^{\mathsf{T}}$ и ($k \times k$)-матрицей точности B (или так называемое $\bar{t}(\nu|\Theta_0\,;B)$-распределение)

Стандартный k-мерный з.р.в. Стьюдента с ν ст. св. описывает распределение k-мерной случайной величины

$$\bar{t}(\nu) = (t^{(1)}(\nu), t^{(2)}(\nu), \ldots, t^{(k)}(\nu))^{\mathsf{T}}, \tag{П.5}$$

где каждая из компонент $t_j(\nu)$ — стандартная стьюдентовская случайная величина (П.1), и все компоненты $t_j(\nu)$ ($j = 1, 2, \ldots, k$) взаимнонекоррелированы. Функция плотности вероятности $f_{\bar{t}(\nu)}(X)$ в точке $X = (x^{(1)}, x^{(2)}, \ldots, x^{(k)})^{\mathsf{T}}$ задается соотношением

$$f_{\bar{t}(\nu)}(X) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{\frac{k}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{1}{\nu} X^{\mathsf{T}} \cdot X\right)^{-\frac{\nu+k}{2}}, \tag{П.6}$$

причем $\mathbf{E}\bar{t}(\nu) = \mathbf{0}_k$ и ковариационная матрица $\Sigma_{\bar{t}(\nu)} = \dfrac{\nu}{\nu-2} \cdot \mathbf{I}_k$, где $\mathbf{0}_k$ обозначает k-мерный вектор-столбец из нулей, а $\mathbf{I}_k$ — единичная матрица размерности k.

Обобщенное k-мерное распределение Стьюдента (или $\bar{t}(\nu|\Theta_0\,;B)$-*распределение*) *с ν ст. св., параметром сдвига Θ_0 и матрицей точности B* описывает распределение случайной величины

$$\bar{t}(\nu|\Theta_0\,;B) = C \cdot \bar{t}(\nu) + \Theta_0, \tag{П.7}$$

где C — некоторая невырожденная $k \times k$-матрица, $B = (CC^{\mathsf{T}})^{-1}$, $\Theta_0 = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)$, а $\bar{t}(\nu)$ — стандартная стьюдентовская k-мерная случайная величина (П.5), подчиняющаяся з.р.в. с плотностью (П.6). Значение функции плотности вероятности $f_{\bar{t}(\nu|\Theta_0;B)}(X)$ случайной величины $\bar{t}(\nu|\Theta_0\,;B)$ в точке $X = (x^{(1)}, x^{(2)}, \ldots, x^{(k)})^{\mathsf{T}}$ задается соотношением

$$f_{\bar{t}(\nu|\Theta_0;B)}(X) = \frac{\Gamma\left(\frac{\nu+k}{2}\right)}{(\pi\nu)^{\frac{k}{2}} \Gamma\left(\frac{\nu}{2}\right)} \cdot |B|^{\frac{1}{2}} \left(1 + \frac{1}{\nu}(X - \Theta_0)^{\mathsf{T}} B (X - \Theta_0)\right)^{-\frac{\nu+k}{2}}, \tag{П.8}$$

причем $\mathbf{E}\bar{t}(\nu|\Theta_0\,;B) = \Theta_0$ и ковариационная матрица $\Sigma_{\bar{t}(\nu|\Theta_0;B)} = \dfrac{\nu}{\nu-2} \cdot B$.

Именно этим з.р.в. описывается сопряженное априорное (а следовательно, и апостериорное) частное распределение вектора $\Theta = (\theta_0, \theta_1, \ldots, \theta_k)^{\mathsf{T}}$ коэффициентов регрессии в нормальной КЛММР, а также условное (при фиксированных $\mathbf{X}$, Y и $\widetilde{\mathbf{X}}$) апостериорное распределение вектора $\widetilde{Y} = (y_{n+1}, y_{n+2}, \ldots, y_{n+q})^{\mathsf{T}}$ прогнозных значений зависимой переменной в этой модели (см. в тексте консультации пример 9 и задачу 5).

При построении байесовских интервальных оценок и доверительных областей для параметров Θ, так же как и при построении байесовских интервальных оценок и доверительных областей для прогнозных значений $\widetilde{Y}$, используются следующие свойства $\bar{t}(\nu|\Theta_0;B)$-распределения.

Свойство (А). Пусть анализируемая k-мерная случайная величина $t(\nu|\Theta_0;B)$ разбита на два подвектора $\bar{t}^{(1)}(\nu|\Theta_0;B)$ и $\bar{t}^{(2)}(\nu|\Theta_0;B)$ соответственно размерностей k_1 и k_2 ($k_1+k_2=k$), т. е.

$$\bar{t}(\nu|\Theta_0;B)=\begin{pmatrix}\bar{t}^{(1)}(\nu|\Theta_0;B)\\ \bar{t}^{(2)}(\nu|\Theta_0;B)\end{pmatrix}.$$

Соответственно этому разобьются на блоки вектор средних значений Θ_0 и матрица точности B:

$$\Theta_0=\begin{pmatrix}\Theta_1(1)\\ \Theta_1(2)\end{pmatrix} \text{ и } B=\begin{pmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{pmatrix}.$$

Сформулируем свойство (А).

Частное (маржинальное) распределение вектора $\bar{t}^{(1)}(\nu|\Theta_0;B)$ *является* k_1*-мерным обобщенным распределением Стьюдента* $\bar{t}(\nu|\Theta_0(1);B(1))$ *с параметром сдвига* $\Theta_0(1)=(\theta_1^0,\ldots,\theta_{k_1}^0)^{\mathrm{T}}$ *и матрицей точности* $B(1)=B_{11}-B_{12}\,B_{22}^{-1}\,B_{21}$.

При построении интервальных оценок нас интересует частное распределение отдельной (j-й) компоненты $t^{(j)}(\nu|\Theta_0;B)$ анализируемой k-мерной обобщенной стьюдентовской случайной величины $\bar{t}(\nu|\Theta_0;B)$. Соответственно, при этом используется частный случай данного свойства, когда в роли $\bar{t}^{(1)}(\nu|\Theta_0;B)$ выступает компонента $t^{(j)}(\nu|\Theta_0;B)$. Тогда:

$$k_1=1;\ \Theta_0(1)=\theta_j^0;\ B(1)=b_{jj}-B_{j\cdot}\cdot B(j)\cdot B_{\cdot j}, \tag{П.9}$$

где b_{jj} — j-й диагональный элемент матрицы точности B, $B_{j\cdot}=(b_{j1},\ldots,b_{j\cdot j-1},b_{j\cdot j+1},\ldots,b_{jk})$ — $(k-1)$-мерная строка, $B_{\cdot j}=(b_{1j},\ldots,b_{j-1\cdot j},b_{j+1\cdot j},\ldots,b_{kj})^{\mathrm{T}}$ — $(k-1)$-мерный столбец матрицы $B(j)$, а $B(j)$ — это $(k-1)\times(k-1)$-матрица, получающаяся из матрицы B вычеркиванием из нее j-й строки и j-го столбца.

Заметим, что в данном случае $B(1)$ — это *числовой параметр точности* в частном обобщенном одномерном стьюдентовском распределении компоненты $t^{(j)}(\nu|\Theta_0;B)$.

Свойство (В) используется при построении *доверительных областей* для неизвестных значений параметров КЛММР или одновременно для нескольких прогнозных значений зависимой переменной и *заключается в том, что статистика*

$$\gamma=\frac{1}{k}\left(\bar{t}(\nu|\Theta_0;B)-\Theta_0\right)^{\mathrm{T}}B\left(\bar{t}(\nu|\Theta_0;B)-\Theta_0\right)$$

асимптотически (по $\nu\to\infty$*) подчиняется* $F(k;\nu)$*-распределению.*

Поэтому, определяя из таблиц по заданной доверительной вероятности P_0 значение 100(1-P_0)%-ной точки $F_{1-P_0}(k;\nu)$ соответствующего F-распределения, мы можем с помощью неравенства

$$\frac{1}{k}\left(\bar{t}(\nu|\Theta_0;B)-\Theta_0\right)^{\mathrm{T}} B\left(\bar{t}(\nu|\Theta_0;B)-\Theta_0\right) < F_{1-P_0}(k;\nu) \qquad (\text{П.10})$$

определить k-мерную область, в которую попадает $100P_0$ % наблюдений случайной величины $\bar{t}(\nu|\Theta_0;B)$.

Приложение 2а

Двумерное гамма-нормальное распределение и его свойства

Совместное двумерное распределение случайной величины $(\theta;h)$ называется **гамма-нормальным**, если его функция плотности вероятности $p(\theta;h)$ задается (с точностью до нормирующего множителя) соотношением

$$p(\theta;\, h) \sim (\lambda_0 h)^{\frac{1}{2}} e^{-\frac{\lambda_0 h}{2}(\theta-\theta_0)^2} \cdot h^{\alpha-1}e^{-\beta h}, \qquad (\text{П.11})$$

где λ_0, θ_0, α и β — некоторые числовые значения параметров этого семейства распределений.

Свойства двумерного гамма-нормального распределения

(I) *Частное распределение параметра* θ есть одномерное обобщенное распределение Стьюдента (П.4) с 2α ст. св., параметром сдвига θ_0 и параметром точности $c = \lambda_0 \cdot \alpha/\beta$, т. е.

$$\theta = t\left(2\alpha|\theta_0;\, \lambda_0\frac{\alpha}{\beta}\right).$$

Отсюда, в частности, следует, что случайная величина $\sqrt{\lambda_0 \dfrac{\alpha}{\beta}}(\theta-\theta_0)$ подчиняется стандартному распределению Стьюдента с 2α ст. св.

(II) *Частное распределение параметра* h есть гамма-распределение (23) с параметрами (α,β) и, следовательно, $\mathbf{E}h = \dfrac{\alpha}{\beta}$, $\mathbf{D}h = \dfrac{\alpha}{\beta^2}$ и $h \in [\gamma_{1-\varepsilon}(\alpha;\beta);\gamma_\varepsilon(\alpha;\beta)]$ с вероятностью $P_0 = 1-2\varepsilon$, где $\gamma_q(\alpha,\beta)$ — это $100q$%-ная точка гамма-распределения с параметрами α и β.

Отметим, что при α, кратном 0,5, справедлива формула:

$$\gamma_q(\alpha,\beta) = \frac{1}{2\beta}\chi_q^2(2\alpha), \qquad (\text{П.12})$$

где $\chi_q^2(m)$ — это $100q$%-ная точка «хи-квадрат»-распределения с m ст. св.

(III) *Условное распределение параметра* θ (при условии заданности значения параметра h, т. е. при $h = h_0$, где h_0 — заданное число) является $(\theta_0;\, 1/\lambda_0 h_0)$-нормальным распределением (вытекает из (П.11) при подстановке в правую часть этого соотношения заданного значения $h = h_0$).

Приложение 2б

Многомерное ($k+1$-мерное, $k>1$) гамма-нормальное распределение и его свойства

Совместное $k+1$-мерное распределение параметров $\Theta = (\theta_1,\theta_2,\ldots,\theta_k)^{\mathrm{T}}$ и h называется **многомерным гамма-нормальным**, если его функция плотности вероятности $p(\Theta;h)$ задается (с точностью до нормирующего множителя) соотношением:

$$p(\Theta;h) \sim h^{\frac{k}{2}}|\Lambda_0|^{\frac{1}{2}} e^{-\frac{h}{2}(\Theta-\Theta_0)^{\mathrm{T}}\Lambda_0(\Theta-\Theta_0)} \cdot h^{\alpha-1}e^{-\beta h}, \qquad (\text{П.13})$$

где заданные численные значения векторного параметра сдвига $\Theta_0 = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)^{\mathrm{T}}$, элементов $(k \times k)$-матрицы точности Λ_0, а также параметров α и β однозначно определяют з.р.в. параметров Θ и h.

Свойства многомерного гамма-нормального распределения

(I) *Частное распределение векторного параметра* $\Theta = (\theta_1, \theta_2, \dots, \theta_k)^{\mathrm{T}}$ есть многомерное обобщенное распределение Стьюдента (П.8) с 2α ст. св., параметром сдвига $\Theta = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)^{\mathrm{T}}$ и матрицей точности $B = \frac{\alpha}{\beta} \cdot \Lambda_0$.

(II) *Частное распределение скалярного параметра* h есть гамма-распределение (23) с параметрами α и β.

(III) *Условное распределение векторного параметра* Θ (при условии заданности значения параметра h, т. е. при $h = h_0$, где h_0 — заданное значение) является k-мерным $(\Theta_0; (h_0 \Lambda_0)^{-1})$-нормальным распределением.

Приложение 3

Некоторые сведения об априорных з.р.в., сопряженных по отношению к наблюдаемым генеральным совокупностям, зависящим от единственного неизвестного параметра

№ пп	З.р.в. наблюдаемой генеральной совокупности	Сопряженный априорный з.р.в. $p(\theta)$, выражения для $\mathbf{E}\theta$ и $\mathbf{D}\theta$	Апостериорный з.р.в. $p(\theta \mid x_1, x_2, \dots, x_n)$, выражения для его параметров
1	$(\theta;\sigma^2)$-нормальный, $f(x\vert\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\theta)^2}{2\sigma^2}}$ (значение дисперсии σ^2 известно)	$(\theta_0;\sigma_0^2)$-нормальный; $\mathbf{E}\theta = \theta_0$; $\mathbf{D}\theta = \sigma_0^2$ (θ_0 и σ_0^2 — заданы)	$(\theta_0';\sigma_0'^2)$-нормальный, где $\theta_0' = \frac{\bar{x} + \gamma\theta_0}{1+\gamma}$ и $\sigma_0'^2 = \sigma^2 / n(1+\gamma)$, а $\gamma = \sigma^2 / n\sigma_0^2$
2	Экспоненциальный $f(x\vert\theta) = \begin{cases} \theta e^{-\theta x} & \text{при } x \ge 0 \\ 0 & \text{при } x < 0 \end{cases}$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$ $(\theta > 0)$ — гамма-распределение; $\mathbf{E}\theta = \alpha / \beta$; $\mathbf{D}\theta = \alpha / \beta^2$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha' = \alpha + n$, $\beta' = \beta + \sum_{i=1}^{n} x_i$
3	$[0;\theta]$-равномерный: $f(x\vert\theta) = \begin{cases} \frac{1}{\theta} & \text{для } 0 \le x \le \theta \\ 0 & \text{для } x \notin [0;\theta] \end{cases}$	$p(\theta) = \begin{cases} \frac{\alpha\theta_0^\alpha}{\theta^{\alpha+1}} & \text{при } x \ge 0 \\ 0 & \text{при } x < 0 \end{cases}$; $(\theta > 0)$ — распределение Парето; $\mathbf{E}\theta = \frac{\alpha\theta_0}{\alpha - 1}$; $\mathbf{D}\theta = \frac{\alpha\theta_0^2}{(\alpha-1)^2(\alpha-2)}$ (α и θ_0 — заданы)	Распределение Парето с параметрами $\alpha' = \alpha + n$, $\theta_0' = \max\{\theta_0; x_1, x_2, \dots, x_n\}$
4	Распределение Пуассона: $P\{\xi = x\} = \frac{\theta^x}{x!} e^{-\theta}$, $x = 0, 1, 2, \dots$	$p(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$ $(\theta > 0)$ — гамма-распределение; $\mathbf{E}\theta = \frac{\alpha}{\beta}$; $\mathbf{D}\theta = \frac{\alpha}{\beta^2}$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha' = \alpha + \sum_{i=1}^{n} x_i$, $\beta' = \beta + n$

Окончание

№ пп	З.р.в. наблюдаемой генеральной совокупности	Сопряженный априорный з.р.в. $p(\theta)$, выражения для $\mathbf{E}\theta$ и $\mathbf{D}\theta$	Апостериорный з.р.в. $p(\theta \mid x_1, x_2, \ldots, x_n)$, выражения для его параметров
5	Биномиальное распределение: $P\{\xi = x\} = C_N^x\, \theta^x (1-\theta)^{N-x}$ (значение параметра N известно)	$p(\theta) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ $(0 \le \theta \le 1)$ — бета-распределение; $\mathbf{E}\theta = \dfrac{a}{a+b}$; $\mathbf{D}\theta = \dfrac{ab}{(a+b)^2(a+b+1)}$ (a и b — заданы)	Бета-распределение с параметрами $a' = a + \sum_{i=1}^{n} x_i$, $b' = b + nN - \sum_{i=1}^{n} x_i$
6	Отрицательное биномиальное распределение $P\{\xi = x\} = C_{x-1}^{k-1} \theta^k (1-\theta)^{x-k}$ (значение параметра k известно) $x = k, k+1, \ldots$	$p(\theta) = \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1}$ $(0 \le \theta \le 1)$ — бета-распределение; $\mathbf{E}\theta = \dfrac{a}{a+b}$; $\mathbf{D}\theta = \dfrac{ab}{(a+b)^2(a+b+1)}$ (a и b — заданы)	Бета-распределение с параметрами $a' = a + kn$, $b' = b + \sum_{i=1}^{n} x_i - kn$
7	Распределение Парето $f(x \mid \theta) = \begin{cases} \dfrac{\theta x_0^{\theta}}{x^{\theta+1}} & \text{при } x \ge x_0 \\ 0 & \text{при } x < x_0 \end{cases}$ (значение параметра x_0 — известно)	$p(\theta) = \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}$ $(\theta > 0)$ — гамма-распределение; $\mathbf{E}\theta = \dfrac{\alpha}{\beta}$; $\mathbf{D}\theta = \dfrac{\alpha}{\beta^2}$ (α и β — заданы)	Гамма-распределение с параметрами $\alpha' = \alpha + n$, $\beta' = \beta + n \ln\left(\dfrac{g_n}{x_0}\right)$, где $g_n = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$

Список литературы

Айвазян С. А. (2001). Прикладная статистика и основы эконометрики. Том 2: Основы эконометрики. Издание 2-е. Юнити. §§ 2.1–2.3.

Айвазян С. А., Мхитарян В. С. (2001а). Прикладная статистика и основы эконометрики. Том 1: Теория вероятностей и прикладная статистика. Издание 2-е. Юнити. § 7.6.

Айвазян С. А., Мхитарян В. С. (2001б). Прикладная статистика в задачах и упражнениях. М.: Юнити.

Де Гроот М. (1974). Оптимальные статистические решения. Пер. с англ. М.: Мир. Гл. 4, 5, 9 и §§ 11.10–11.12.

Зельнер А. (1980). Байесовские методы в эконометрике. Пер. с англ. М.: Статистика. Главы 1–3.

Ghosh J. K., Delampady M., Samanta T. (2006). An Introduction to Bayesian Analysis. Theory and Methods. Springer.

Jeffreys H. (1957). Scientific Inference. 2nd ed. Cambridge University Press.

Lancaster A. (2004). An Introduction to Modern Bayesian Econometrics. Blackwell Publ.